

AI TOOLS, NOT GODS

AI TOOLS, NOT GODS

WHY ARTIFICIAL INTELLIGENCE
HYPE THREATENS GLOBAL
GOVERNANCE—AND HOW TO FIX IT

Caroline De Cock

BTF Press | Antwerp, Belgium

9789083616803 (paperback)

9789083616810 (eBook)

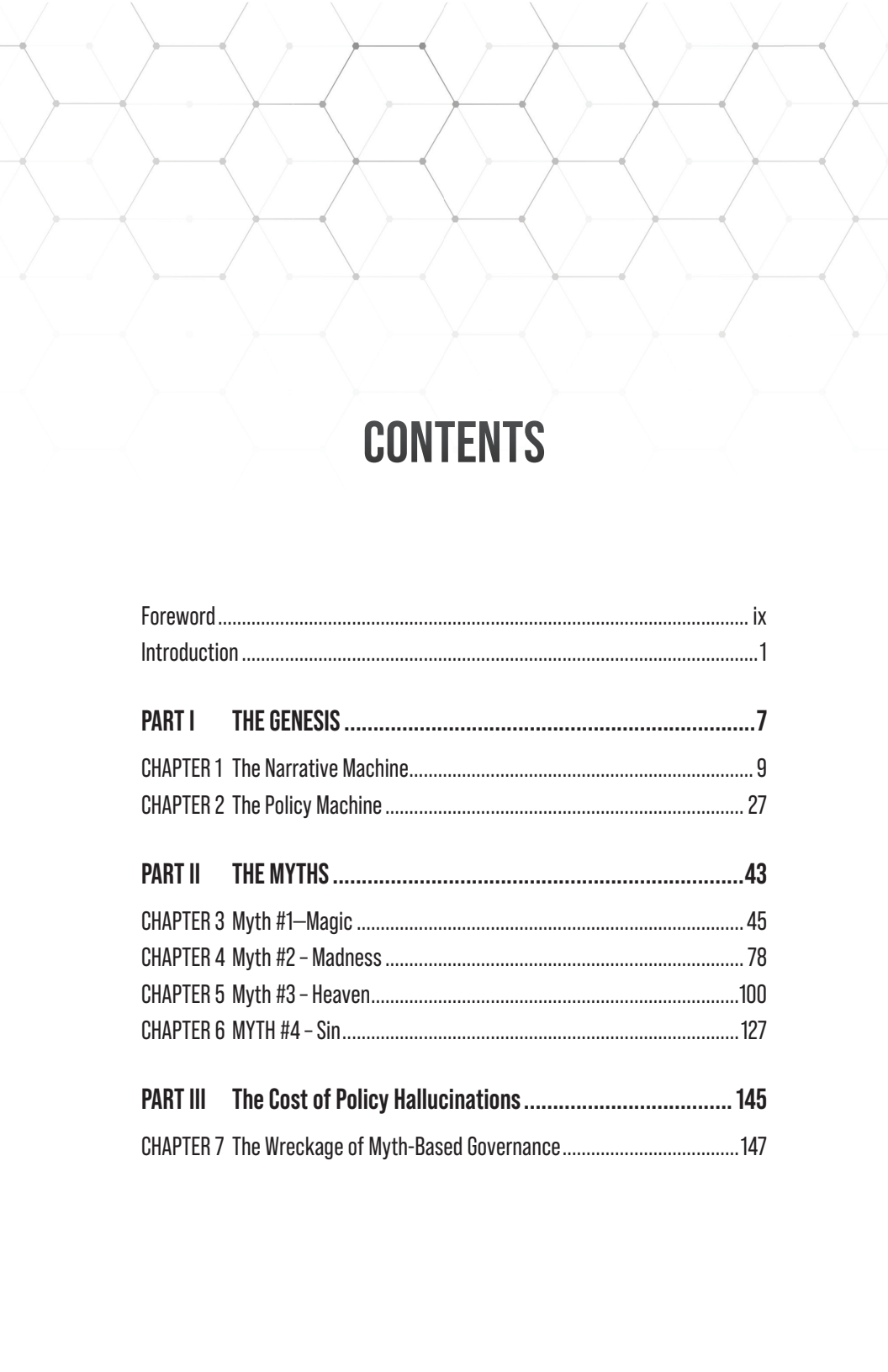
Published by BTF Press, 2026

publishing@btfpress.org

This work is licensed under a Creative Commons CC0 1.0
Universal (CC0 1.0) Public Domain Dedication license.

To my family.

And to all the chatty, sycophantic bots that supported my book-writing journey by mindlessly telling me how amazing I am. Your enthusiasm was unbelievable. Literally unbelievable.



CONTENTS

Foreword	ix
Introduction	1
 PART I THE GENESIS	 7
CHAPTER 1 The Narrative Machine.....	9
CHAPTER 2 The Policy Machine	27
 PART II THE MYTHS	 43
CHAPTER 3 Myth #1—Magic	45
CHAPTER 4 Myth #2 – Madness	78
CHAPTER 5 Myth #3 – Heaven.....	100
CHAPTER 6 MYTH #4 – Sin.....	127
 PART III The Cost of Policy Hallucinations	 145
CHAPTER 7 The Wreckage of Myth-Based Governance.....	147

PART IV A Blueprint for AI Governance 167

CHAPTER 8 The Foundation: from Graveyard to Blueprint169

CHAPTER 9 Principles for Mature AI Governance.....180

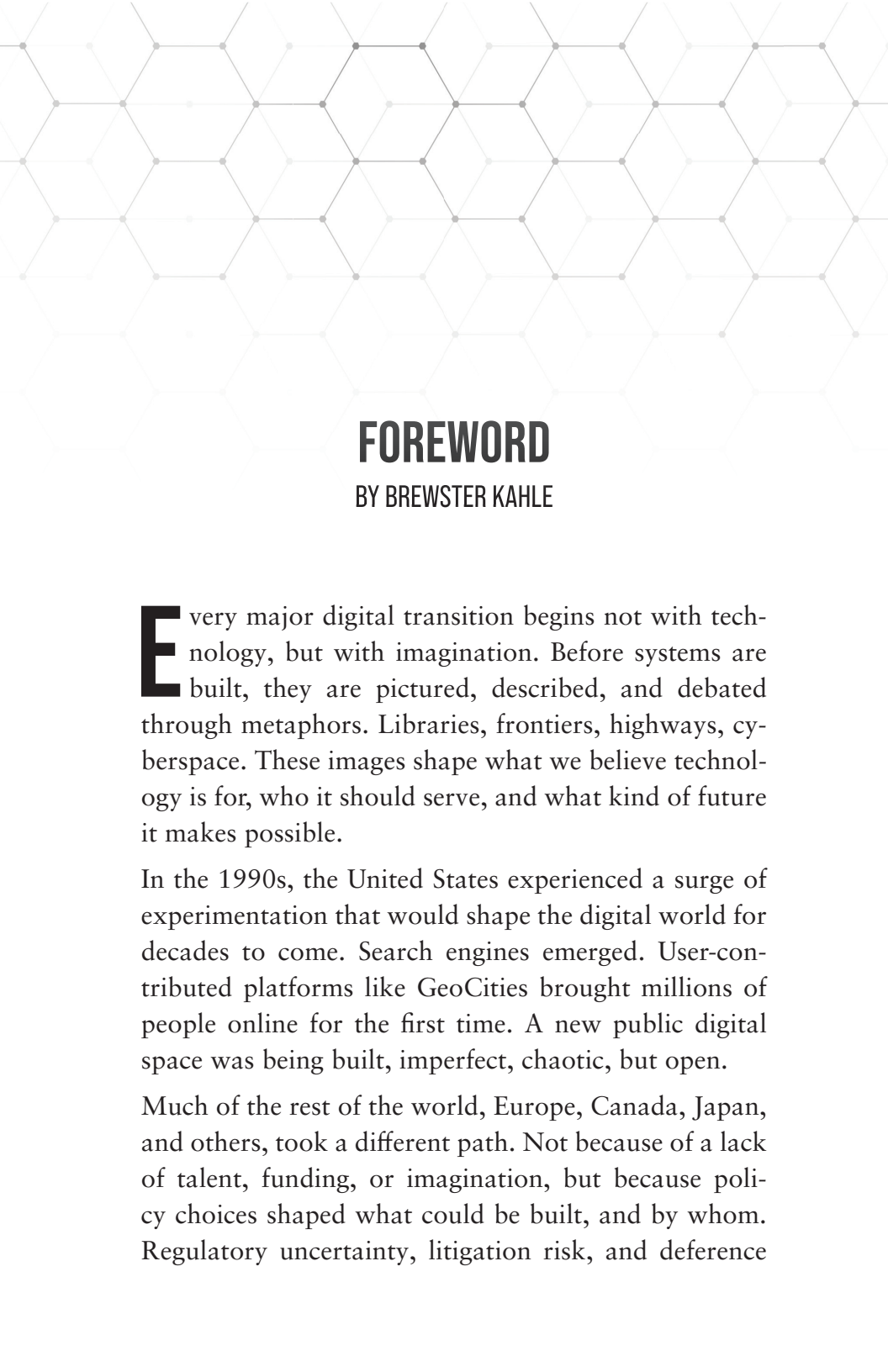
Epilogue.....215

Endnotes219

Bibliography271

Acknowledgments319

About the Author321



FOREWORD

BY BREWSTER KAHLE

Every major digital transition begins not with technology, but with imagination. Before systems are built, they are pictured, described, and debated through metaphors. Libraries, frontiers, highways, cyberspace. These images shape what we believe technology is for, who it should serve, and what kind of future it makes possible.

In the 1990s, the United States experienced a surge of experimentation that would shape the digital world for decades to come. Search engines emerged. User-contributed platforms like GeoCities brought millions of people online for the first time. A new public digital space was being built, imperfect, chaotic, but open.

Much of the rest of the world, Europe, Canada, Japan, and others, took a different path. Not because of a lack of talent, funding, or imagination, but because policy choices shaped what could be built, and by whom. Regulatory uncertainty, litigation risk, and deference

to incumbent interests made large-scale experimentation harder. Entrepreneurs hesitated. Investors looked elsewhere. Over time, entire categories of public-facing digital infrastructure simply did not emerge.

Policy matters, not because it predicts innovation, but because it creates the conditions under which innovation can happen.

Today, generative AI offers another inflection point. For the first time in years, the dominance of search and information intermediaries is no longer inevitable. New tools make it possible to imagine systems that are more plural, more local, more publicly oriented, and better aligned with education, research, and cultural access.

Some jurisdictions are beginning to act with that future in mind. Japan, for example, has chosen a clear and pragmatic approach, allowing AI systems to learn from copyrighted works while holding outputs to account. In other words, enable learning, evaluate results. This creates space for innovation while preserving accountability, a balance that invites builders to build.

Europe stands at a similar crossroads. It has world-class researchers, strong public institutions, and a deep commitment to cultural heritage and education. It also has a unique opportunity to design AI policy that supports openness, competition, and democratic access to knowledge. Narrow exemptions and hesitant signals, however well-intentioned, are not enough to unlock that potential, especially when institutions tasked with preserving culture remain unsure whether they are truly allowed to use the tools now available to them.

History offers lessons here. Canada once developed OpenText, an early and promising search technology. Rather than scaling it into a public-facing infrastructure, it retreated into safer, narrower applications, and the opportunity passed. Europe can choose differently this time.

The question is not whether generative AI should exist, but what kind of information ecosystem we want it to support. One shaped primarily by incumbency and caution. Or one that empowers libraries, schools, museums, researchers, startups, and citizens to participate fully in the digital transition.

No one knows these battles better than Caroline De Cock, who has spent decades pursuing an open web in Brussels, the European Union's centralized location for these regulations. The public interests embodied in our schools, libraries, and museums exist to expand access to knowledge and to help societies learn, especially in moments of technological change. With thoughtful, confident policy design, those institutions can help ensure that AI strengthens, rather than narrows, our shared intellectual commons.

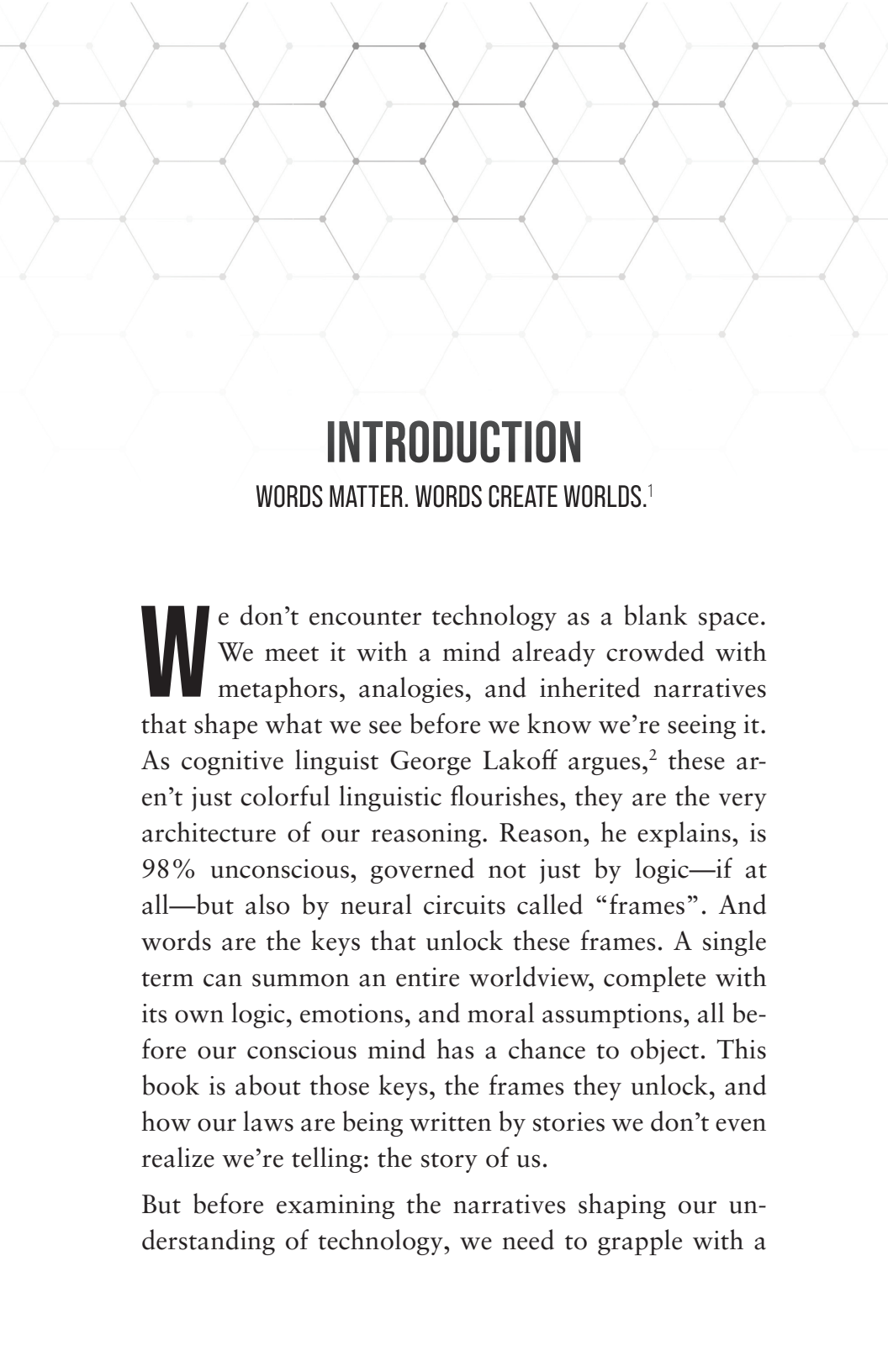
Europe has an opportunity not just to regulate generative AI, but to imagine what it could become, a tool for learning, creativity, resilience, and democratic participation. The choices made now will shape its cultural, economic, and security landscape for years to come.

Policy matters.

Brewster Kahle

Founder of the Internet Archive

Internet Hall of Fame Inductee



INTRODUCTION

WORDS MATTER. WORDS CREATE WORLDS.¹

We don't encounter technology as a blank space. We meet it with a mind already crowded with metaphors, analogies, and inherited narratives that shape what we see before we know we're seeing it. As cognitive linguist George Lakoff argues,² these aren't just colorful linguistic flourishes, they are the very architecture of our reasoning. Reason, he explains, is 98% unconscious, governed not just by logic—if at all—but also by neural circuits called “frames”. And words are the keys that unlock these frames. A single term can summon an entire worldview, complete with its own logic, emotions, and moral assumptions, all before our conscious mind has a chance to object. This book is about those keys, the frames they unlock, and how our laws are being written by stories we don't even realize we're telling: the story of us.

But before examining the narratives shaping our understanding of technology, we need to grapple with a

basic linguistic question: what do we mean when we say “Artificial Intelligence?” The term “AI” is routinely used to describe a dizzying variety of technologies (generative models, statistical classifiers, recommender systems, predictive analytics, and more) that share little in common beyond the vague promise of computational cleverness. It’s basically a branding label masquerading as a technical definition.

In fact, as U.S. journalist Karen Hao documents in *Empire of AI*, the term itself was a marketing tool from its inception.³ When U.S. cognitive scientist and one of the founders of AI, John McCarthy, convened the seminal Dartmouth Summer Research Project on Artificial Intelligence in 1956, he swapped the less exciting term “automata studies” for “artificial intelligence” precisely because it was more evocative and better at attracting interest and funding.⁴

Using “AI” as an umbrella term amounts to calling everything with buttons a “controller”, whether it’s a TV remote, a pacemaker programmer, or a nuclear launch panel, and then trying to draft legislation on the safety and social impact of said controllers. The media, ever helpful, would leap into the void with headlines like “Controllers: Existential Threat or Next-Gen Convenience?” Talk shows would stage theatrical debates where one guest warns of imminent nuclear annihilation, another enthuses about ad-skipping convenience during *The Great British Bake Off*, and a third attempts (futilely) to explain that their “controller” just helps manage heart arrhythmias in elderly patients. As a result of these supposedly nuanced and fact-based

debates, public understanding would swing wildly between gadget fetishism and apocalyptic dread, depending entirely on which controller a person had in mind.

As the authors of *AI Snake Oil*, which aims to debunk the AI hype, observe, most of the true “snake oil” is concentrated in the realm of predictive AI: systems that claim to make high-stakes judgments about people’s lives, from loan applications to hiring decisions.⁵ These tools are often sold on the promise of superhuman foresight, a claim that rarely withstands scrutiny. Generative AI, like the large language models (LLMs)⁶ that have captured the world’s attention, is different. Its progress is remarkable, but as a product, it is still immature and unreliable, often functioning as what some have aptly called a “bullshitter”—a system designed to produce plausible text rather than necessarily true statements.

We fall into the trap of debating AI as a single entity when we should be having different conversations about distinct tools. However, this is precisely what happens when we bundle ChatGPT with facial recognition, military drones, and supply chain algorithms under a single, mythically loaded two-letter acronym. The ensuing discourse is so broad it often collapses into caricature. Grouping together vastly different applications under one label isn’t just analytically sloppy, it’s politically dangerous. It trivializes genuine risks, demonizes useful tools, and traps policymakers in a tug-of-war between sci-fi panic and PR-driven techno-optimism. Yet that is the AI discourse picked up by media and policy.

Such a terminological mess matters, not just because it creates confusion, but because it actively shapes how we think about the risks, benefits, and governance of these technologies. So before we talk policy, we need to talk about the stories that shape our headlines and our laws. Because AI isn't just built by developers. What we recognize as AI is being shaped through four powerful myths that inform how we see it, sell it, and surrender power to it. We "story" AI into being. Metaphors become policies, myths become institutionalized, and democratic agency erodes slowly when governance is built on ill-informed fantasy rather than facts. Hence, "AI" is more than the buzzword *du jour*. How we conceive of AI is important because its rapid evolution is the ultimate stress test for our traditional methods of governance.

AI is the focus of this book because it serves as the perfect contemporary case study, a crucible in which our old, technology-specific regulatory habits are being tested and found wanting. The patterns of hype, fear, and institutional capture we see today echo the histories of the internet, nuclear power, and many other technologically driven or facilitated shifts. By dissecting the myths surrounding AI, we can forge the durable, technology-neutral principles needed to govern whatever comes next.

In **Part I: The Genesis**, we trace the origins of these myths, from the "Wild West" of the early internet to President Eisenhower's glowing faith in "Atoms for Peace." We then examine the policy machine itself—the institutional pipeline of academia, media, and politics

that translates these compelling stories into concrete laws.

Part II: The Myths, dissects the four great illusions governing AI. We explore the myth of Magic, which cloaks technology in wonder to deflect accountability; the myth of Madness, which swings between manic gold rushes and paralyzing moral panics; the myth of Heaven, which promises salvation from our human failings; and the myth of Sin, which casts the algorithm as a digital scapegoat when that salvation fails.

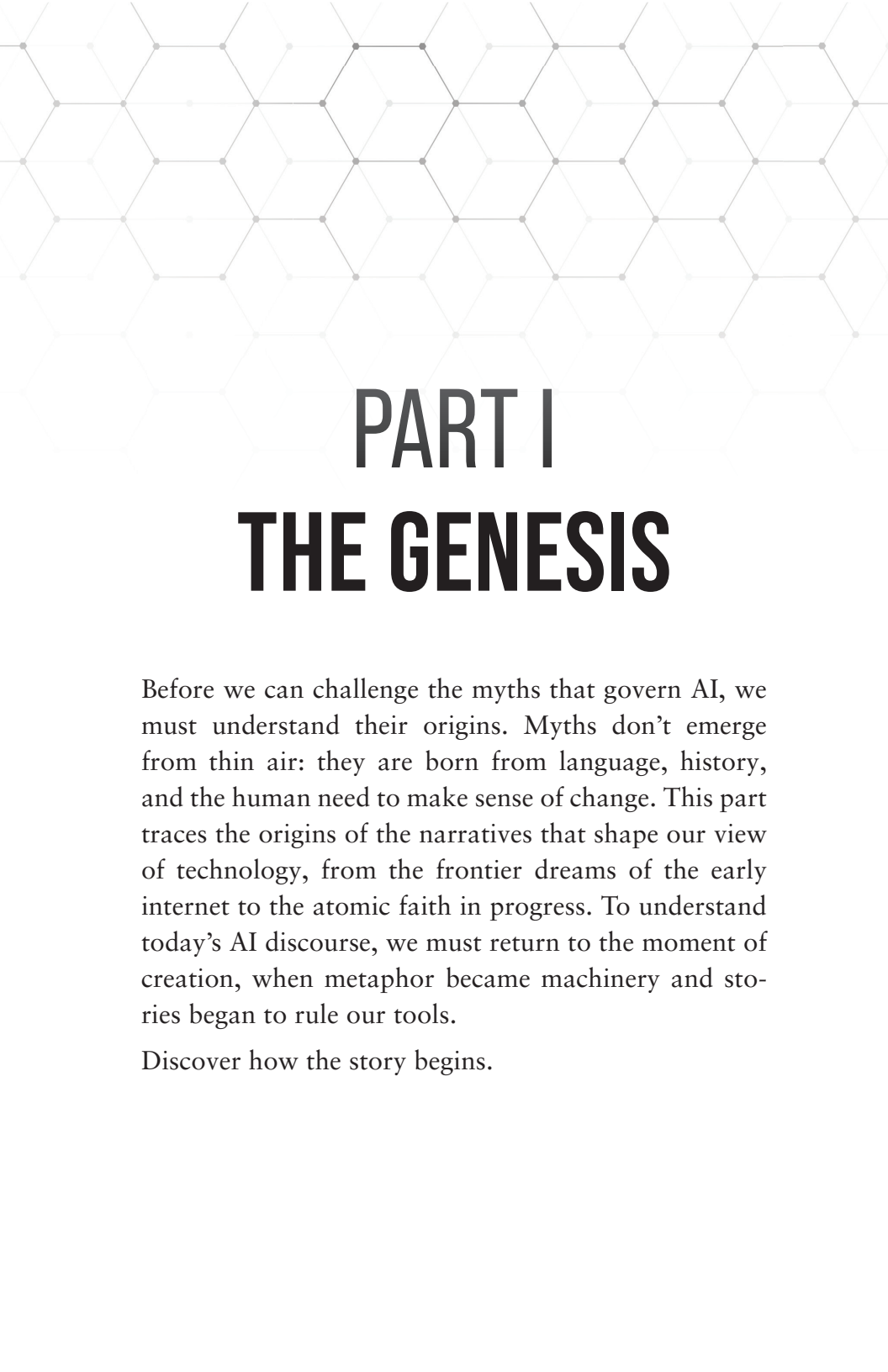
Part III: The Cost of Policy Hallucinations, takes stock of the real-world damage. When governance is built on myth, we suffer predictable failures: policy distraction that chases phantoms, governance whiplash from knee-jerk reactions, and the erosion of democratic control as we cede accountability to prophets and bots.

Part IV: A Blueprint for AI Governance, provides the antidote. It unbuilds what doesn't serve us, moving beyond the graveyard of obsolete, technology-specific laws to set out a practical, technology-neutral blueprint built on three principles: Accountability, Proportionality, and Durability.

Finally, the **Epilogue** calls for an "Age of Unenchantment," urging us to reject technological determinism and embrace the patient, necessary work of democratic governance.

Taken together, these explorations show that technology is never just engineered, it's narrated into being. If we want governance rooted in reason rather than illusion, we must begin by choosing our words as carefully

as our tools. Because before we can write better laws,
we need to tell better stories.

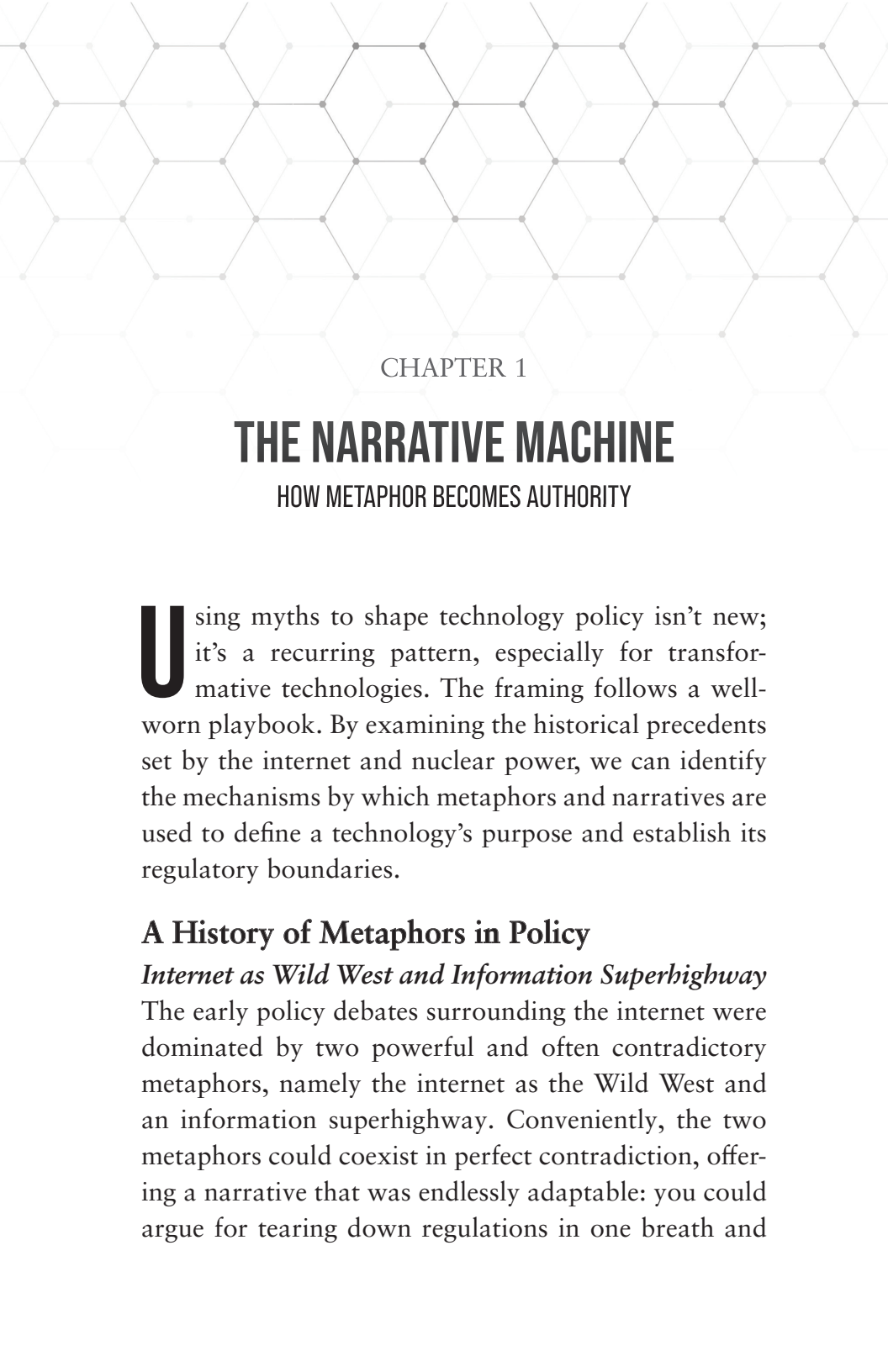


PART I

THE GENESIS

Before we can challenge the myths that govern AI, we must understand their origins. Myths don't emerge from thin air: they are born from language, history, and the human need to make sense of change. This part traces the origins of the narratives that shape our view of technology, from the frontier dreams of the early internet to the atomic faith in progress. To understand today's AI discourse, we must return to the moment of creation, when metaphor became machinery and stories began to rule our tools.

Discover how the story begins.



CHAPTER 1

THE NARRATIVE MACHINE

HOW METAPHOR BECOMES AUTHORITY

Using myths to shape technology policy isn't new; it's a recurring pattern, especially for transformative technologies. The framing follows a well-worn playbook. By examining the historical precedents set by the internet and nuclear power, we can identify the mechanisms by which metaphors and narratives are used to define a technology's purpose and establish its regulatory boundaries.

A History of Metaphors in Policy

Internet as Wild West and Information Superhighway

The early policy debates surrounding the internet were dominated by two powerful and often contradictory metaphors, namely the internet as the Wild West and an information superhighway. Conveniently, the two metaphors could coexist in perfect contradiction, offering a narrative that was endlessly adaptable: you could argue for tearing down regulations in one breath and

for building digital tollbooths in the next, all without missing a step.

The Wild West was a political metaphor that could ride in any direction you wanted, and it did. For libertarians, tech founders, and assorted digital cowboys, it represented a wide-open frontier promising endless opportunity with no bureaucrats in sight. It offered a system based on spontaneous order whereby communities could set their own rules, stake their digital claims, and make their fortune without a sheriff's badge flashing into view. They saw themselves as the miners and cattlemen of cyberspace prospecting and building homesteads as they exploited the potential of the internet. This version of what the internet represented was rolled out whenever someone wanted to head off regulation, whether it pertained to content moderation, online marketplaces, or financial oversight.¹

But for regulators, the Wild West wasn't a land of opportunity, it was a saloon brawl on repeat—a lawless frontier demanding the strong hand of a sheriff. Cue calls for surveillance, stricter rules, and more legal power.

That is where the political beauty of the metaphor lay: you could use it to justify doing absolutely nothing... or doing absolutely everything. The framing did not stop there: it is alive and well today. Just look at how cryptocurrency is branded as the Wild West that must be tamed.² Or if you are into cheesy statements, check out the moment former European Commissioner Thierry Breton introduced the EU's Digital Services Act in 2022 with all the subtlety of a spaghetti western, star-

ring the new piece of legislation as the “new sheriff in town” and referencing a clip from *The Good, the Bad and the Ugly* showing gun-toting bounty hunter Clint Eastwood in a cowboy hat in the final confrontation.³

If the Wild West depicted the lawless frontier, the concept of an Information Superhighway promised the opposite: gleaming asphalt, orderly lanes, and well-posted exits. Popularized by the Clinton–Gore administration in the 1990s, it framed the internet as critical national infrastructure—the digital sibling of the Interstate Highway System.⁴ This was the policy equivalent of swapping cowboy boots for a hard hat.

The highway metaphor put the internet’s economic role front and center. It was all about speed, efficiency, and the movement of “goods,” here meaning packets of information.⁵ It aligned policy neatly with a commerce-first, marketplace model, where regulation was about keeping the traffic flowing, not inspecting what was in the trucks. As one contemporary analysis noted, the emphasis was on “speed and quantity of goods... transmitted, not the quality or content.”⁶ That framing opened the door to massive public and private investment, but it also nudged priorities toward commercial applications, quietly shunting community-building, education, and non-profit uses into the slow lane.⁷

It was a metaphor that also smuggled in some baggage. Highways are linear, centralized, and predictable. The internet, decentralized, messy, and allergic to straight lines, was never going to fit neatly into such a traffic plan. Yet the metaphor conveniently obscured those differences, forcing the network into a familiar regulatory

box designed for dedicated point-to-point transportation and traditional telecommunications systems.⁸ Still, the image was reassuring, familiar, and—like all good political metaphors—it told people exactly what they wanted to hear while leaving plenty unsaid.

The history of internet policy shows that the most influential metaphors are rarely the most technically precise, but rather those that best serve political and economic interests. The Wild West and Information Superhighway may seem to have pulled in opposite directions, but both pushed the network along the same path: away from the non-commercial, community-minded hacker ethic of its early years and toward a system dominated by private enterprise and commercial priorities. They provided the language and framing necessary to transform a publicly funded experiment into a marketplace.

More than a historical curiosity, it is a blueprint for understanding today's debate over AI. The words we choose do more than describe new technologies; they set the boundaries of acceptable policy and embed particular economic and political ideologies into their foundations from the outset.

Atoms for Peace vs. Nuclear Devastation

The introduction of nuclear technology to the world was accompanied by a narrative duality of unprecedented extremity that seems familiar when compared to AI narratives: it was presented simultaneously as the ultimate tool of salvation because of its potential to supply near-limitless energy and as the ultimate weapon of annihilation because of its capacity for destruction. This profound tension between a utopian myth

of boundless energy and a dystopian myth of total destruction shaped decades of energy, military, and international policy. The nuclear case is perhaps the starkest historical example of how policy can become completely captured by powerful, mythological narratives, often leading to outcomes divorced from technical or economic reality.

The utopian vision was masterfully crafted and deployed by President Dwight D. Eisenhower's Atoms for Peace program, announced in his speech to the United Nations in 1953.⁹ This initiative was a well-thought-out act of "narrative engineering", the deliberate crafting of stories that make technological developments sound inevitable, revolutionary, and beneficial, while diverting attention from any possible downsides.¹⁰ In this case, the objective was to reframe the terrifying power of the atom, until then associated only with the bombs dropped on Hiroshima and Nagasaki, as a source of immense good for humanity. The program's central promise was of a future powered by clean, abundant nuclear energy that would be "too cheap to meter."¹¹ The slogan never became reality, but it captured imaginations worldwide. The narrative of "Atoms for Peace" positioned nuclear technology as a tool to alleviate global hunger and poverty, promote economic development, and ultimately ensure peace through shared prosperity. This utopian framing served a clear policy goal: to build public acceptance for a domestic civilian nuclear industry and to justify the massive government investment and legislative changes required to bring it into being. Sounds familiar? That's because those who

advocate for the benefits of AI follow the same playbook.

The counter-narrative, of course, was the ever-present threat of nuclear war. It was fueled by metaphors of apocalypse, inferno, and devastation, which saturated Cold War social discourse.¹² The image of the mushroom cloud became a globally recognized icon of humanity's capacity for self-destruction. The *Bulletin of the Atomic Scientists'* Doomsday Clock, introduced in 1947, served as a powerful and enduring visual metaphor, constantly ticking down the minutes to a nuclear apocalypse and reinforcing a sense of impending doom.¹³ This narrative of existential risk went beyond cultural background noise; it was the foundational logic of Cold War geopolitical strategy. It justified the creation of vast nuclear arsenals, the doctrine of mutually assured destruction (MAD), and a state of permanent military readiness, while also fueling a powerful anti-nuclear social movement that challenged this logic.¹⁴

Despite pulling in opposite directions, these two myths produced a similar political outcome: control over nuclear technology was concentrated in the hands of a small, insulated elite of scientists, military leaders, and senior policymakers. This dynamic, where complexity and high stakes justify sidelining public input, offers a stark warning for today's AI debates. The atom was portrayed as too complex, too powerful, and too dangerous for democratic governance. Whether deployed for peace or for war, it was to be managed by a priesthood of nuclear experts holding the keys to both salvation and damnation. The public's role was reduced to

cheering the dream or fearing the nightmare, but never sharing in decision-making.

The nuclear case study serves as a crucial cautionary tale for AI policymaking. It demonstrates how a technology's governing narrative can become detached from its material reality. The promise of energy "too cheap to meter" was a fantasy that never materialized; in reality, the nuclear industry has been plagued by exorbitant costs, complex regulatory hurdles, and the unsolved problem of waste disposal. Yet, the sheer power of that utopian myth was sufficient to drive policy and investment for decades. This reveals a critical lesson: policy often responds not to a technology's demonstrated performance, but to the power of its promise. The current hype-driven cycle of investment and policy-making in AI, fueled by similarly grand promises of solving all of humanity's problems, follows this exact historical pattern. The extreme duality of the AI discourse as both world-savior and world-destroyer risks replicating the nuclear outcome of a policy landscape divorced from reality and a concentration of power in the hands of a narrow technocratic elite.

Storying AI Into Being

When OpenAI released ChatGPT in November 2022, something remarkable happened that had nothing to do with the technology itself. Within hours of the system's debut, journalists, policymakers, and ordinary users began describing their interactions in strikingly similar terms. They spoke of "conversations" with an artificial "mind," marveled at its apparent "creativity,"

and worried about its capacity for “deception.” Some claimed it demonstrated “understanding,” while others insisted it was frequently “hallucinating.”¹⁵

None of this was technically accurate. ChatGPT and other general-purpose AI models do not think, understand, or converse. They process text and images using statistical pattern matching, optimizing probability distributions across vast datasets to predict the most likely next word or pixel in a sequence.¹⁶ In essence, the system is a masterful autocomplete, a sophisticated parrot with a PhD in probability, but a parrot nonetheless.¹⁷ Yet the language people chose to describe this process revealed something profound about how human societies encounter new technologies: they do not simply observe and analyze them. They story them into existence.

The tales we tell about AI are not innocent descriptions of technological capability. They actively shape how this technology develops, how institutions respond to it, and ultimately, how it transforms society. When we refer to AI systems as “intelligent,” we don’t merely describe their functionality; we invoke expectations, fears, and regulatory frameworks associated with intelligence itself.¹⁸ When we speak about machine “learning” and “training,” we import assumptions about teaching, development, and the relationship between teacher and student that may have little relevance to underlying processes such as gradient descent optimization or model fine-tuning.¹⁹

Contemporary AI discourse is often rooted in mythological frameworks that systematically distort what

these systems are and what they might become. These myths influence public opinion and form the conceptual foundation on which policies are written, investments are made, and governance structures are built.²⁰ Understanding them can be seen as an academic exercise. But it's also an urgent practical necessity for anyone who hopes to navigate the age of algorithmic power wisely, and it is definitely a vital necessity for policymakers preparing for the societal and technological changes ahead.

The Unusual Case of AI

AI presents an unusual case. Most technological metaphors emerge *after* a technology's social impact becomes apparent, serving primarily explanatory functions. We called early automobiles "horseless carriages" because they resembled familiar transportation while clearly differing from it. We described the internet as an Information Superhighway to help people understand its connective potential within existing infrastructure and commerce frameworks.²¹

AI metaphors work differently. Many of our dominant framings predate the technology they purport to describe, emerging from science fiction, mythology, and philosophical speculation about machine consciousness.²² We inherited concepts such as artificial intelligence, machine learning, and neural networks from researchers and storytellers who originally imagined rather than observed technological capabilities.²³ This reversal creates a peculiar dynamic in which reality must measure itself against fiction, and where techni-

cal achievements are consistently interpreted through lenses crafted for entirely different purposes. It's like predicting the future based on expertise gathered by watching *Futurama*.²⁴ We are trying to navigate the streets of a new city using a map drawn by ancient cartographers who only dreamed of its existence.

The foundational terminology has profound implications at multiple levels. The term “artificial intelligence” immediately evokes comparisons with human cognition, suggesting that machines might think, reason, or understand in ways analogous to biological minds. This framing shapes everything from research priorities (how do we make machines more like humans?) to regulatory concerns (what happens when artificial minds surpass human capabilities?). Yet, current AI systems operate through mechanisms that bear little resemblance to human cognitive processes, making the intelligence metaphor more misleading than illuminating.²⁵ Some critics, including acclaimed science fiction author Ted Chiang, have been arguing that “artificial intelligence” was a “poor choice of words” back in the 1950s. He suggests “applied statistics” would be a more precise, albeit less glamorous, descriptor, which could have helped avert much of the confusion we face today.²⁶

The language of machine learning carries similar baggage, importing assumptions about education, development, and knowledge acquisition that may obscure rather than clarify how these systems function.²⁷ When we describe algorithmic training, we evoke images of careful instruction, gradual improvement, and pur-

poseful skill development. In practice, training an AI model involves mathematical optimization procedures that adjust millions or billions of parameters to minimize prediction errors—a process with no intention, comprehension, or curiosity. The metaphor lends a false sense of intentionality and accessibility to what, in reality, is an opaque and often counterintuitive statistical process.

Crucially, these metaphors actively steer its development.²⁸ When researchers pursue “artificial general intelligence,” they are not merely trying to improve computational capabilities, they are attempting to fulfill a specific metaphorical vision of what AI should become. The metaphor does not describe the destination, it creates it.

The Four Myths Governing AI Discourse

A close reading of how AI is described across academic papers, policy briefs, news coverage, and pop culture reveals four dominant mythological frameworks at work.²⁹ These are not just different takes on the same technology, they are entirely different ways of defining what AI is, what it does, and why it matters. They also indicate that selecting a metaphor is an inherently political act, prioritizing certain understandings and potential policy responses while marginalizing others, often without conscious intent.

The four myths are:

Myth #1—Magic

The mythology of wonder and inexplicability. In this framework, AI is presented as possessing near-human

or even supernatural capabilities that defy ordinary technological explanations.³⁰ This narrative works by anthropomorphizing AI systems, encouraging us to see them as thinking entities capable of understanding rather than complex statistical tools. Think of headlines marveling at chatbots understanding users. This sense of enchantment is deliberately cultivated, and the trick works because it hides the scaffolding, namely, the data pipelines, the engineering grind, the thousands of human hands labeling training sets for pennies. Such mystification is politically convenient: for the tech industry, it helps deflect calls for transparency by framing AI as too complex; for politicians, it creates a powerful but poorly defined force they can promise to either champion or rein in, depending on the day.

Myth #2—Madness

The mythology of chaos and uncontrollability. In this frame, AI is cast as an unpredictable force whose inner workings are beyond human comprehension, turning opacity from a fixable engineering flaw into an inherent, almost mystical property of the technology.³¹ It fuels both the AI arms race rhetoric and sudden calls for moratoriums. This serves two purposes: companies can shrug off failures as the inevitable quirks of an inscrutable machine, and policymakers can ride the panic or hype to justify almost anything—from bans that attempt to allay speculative fears to a raft of deregulation and subsidies handed out to avoid falling behind. In this frame, unpredictability is not a warning, it is the justification for charging ahead.

Myth #3—Heaven

The mythology of salvation and transcendence. This framework positions AI as humanity's pathway to solving existential challenges that have defined the human condition throughout history.³² Disease, poverty, environmental degradation, and mortality itself become technical problems awaiting algorithmic solutions. This underlies promises that AI will cure diseases or solve climate change. This is politicians' favorite myth, because it promises decisive action without messy compromise. For industry, it's a moral mission statement and a business plan rolled into one. It produces a rationale for investment on a scale usually reserved for wartime, coupled with pleas for regulation so light that it barely touches the ground.

Myth #4—Sin

The mythology of corruption and damnation. In this framework, AI is cast as a digital scapegoat, the ultimate villain: biased,³³ manipulative, job-killing, and culture-draining. This fuels narratives blaming automation for unemployment or algorithms for societal division. It treats technology as the primary source of complex social ills, allowing society to blame bots for deep-seated human problems. It gives politicians a way to propose simple technological restrictions rather than confronting difficult structural reforms or reevaluating outdated legal constructs. It also lets tech companies deflect systemic criticism by focusing the debate on narrow issues of algorithmic bias, which can be framed as technical bugs to be fixed rather than fundamental flaws in their business models. The sin mythology

treats technological development not as progress but as moral decline, where each algorithmic advance takes humanity further from authentic existence.

What makes these mythologies particularly powerful is their ability to be selectively combined, offering a convenient narrative for any occasion.³⁴ A single policy document might simultaneously invoke AI's magical problem-solving capabilities while warning of its sinful potential for bias and manipulation. A technology company can promise the heavenly benefits of its products while, in the same breath, admitting to the madness of systems too complex for human oversight, thereby justifying both its ambition and its opacity. These are not accidental contradictions; they are the core of the strategy. This narrative flexibility allows powerful actors to frame the technology in whatever way best serves their immediate interests—be it securing investment, deflecting accountability, or justifying a particular policy. Instead of weakening the overall story, this ability to mix and match frameworks ensures that the complex, human realities of power, choice, and responsibility are always kept just out of focus, replaced by a simpler, more convenient fiction.

The Seductive Simplicity of Metaphorical Thinking

Why do these metaphorical frameworks prove so compelling, even among supposedly rational policymakers, truth-seeking journalists, and technical experts? The answer lies in what cognitive scientists refer to as the “complexity gap” between human psychological ca-

pabilities and the technical limitations of contemporary AI systems.³⁵ These technologies operate through mechanisms that are simultaneously too complex for intuitive understanding and too abstract for direct observation. Machine learning involves mathematical processes across parameter spaces with millions or billions of dimensions, optimizing objectives through procedures that can require weeks of computation across thousands of specialized processors.

When faced with such complexity, human cognition naturally reaches for simplifying frameworks that make abstract processes seem more familiar and manageable.³⁶ Mythological thinking provides exactly this kind of cognitive scaffolding, allowing people to feel that they understand technologies whose actual operation exceeds their analytical capabilities. Magic tells us that incomprehensibility is proof of power. Madness reassures us that unpredictability is simply its nature. Heaven promises that all our problems are solvable if we just build the right machine. Sin warns us that the same machine is already corrupting our world.

These simplifying frameworks become particularly attractive during periods of rapid technological change, when both experts and the general public struggle to anticipate the social implications of emerging capabilities.³⁷ Mythological thinking provides emotional and intellectual comfort by suggesting that unprecedented situations can be understood through familiar categories drawn from human history, literature, and cultural memory.

The problem is that mythological simplifications, while psychologically satisfying, often prove systematically misleading when translated into policy frameworks.³⁸ They invite us to govern AI as though it were a wizard, a trickster, a savior, or a sinner—all categories that bear little resemblance to the reality of complex statistical tools with bounded capabilities and concrete limitations. When these myths harden into policy frameworks, we end up regulating to contain imagined dangers or to accelerate imagined miracles, while the real, measurable risks and opportunities go unaddressed. The result is governance that may feel decisive but is misaligned with the technology itself, tackling the shadows of our own stories rather than the reality of the systems that have been built.

Beyond the Metaphor, the Stakes of Clarity

The myths that dominate AI discourse are not harmless shortcuts, they are structural obstacles to effective governance, democratic participation, and genuinely beneficial technological development.³⁹ When we view AI through the lens of myth rather than evidence, we leave ourselves open to two equal and opposite failures: uncritical enthusiasm that treats every breakthrough as salvation, and paralyzing fear that casts every advance as the opening scene of a disaster movie. Either reaction stalls the steady, often unglamorous institutional work that emerging technologies demand.

The metaphors and archetypes we use to frame AI shape where investment flows, which research agendas flourish, how regulators set their priorities, and wheth-

er nations choose to cooperate or compete.⁴⁰ They shape public opinion about which uses of AI deserve encouragement and which deserve prohibition. Most importantly, they determine whether societies will build the institutional capacity to govern algorithmic systems or drift into a future where decisions are dictated by forces we barely understand and cannot control democratically.

The way forward is not to strip the wonder from technological achievement, nor to dismiss legitimate concerns about algorithmic power. It is to replace seductive fictions with clearer, more grounded frameworks for understanding what AI is, how it works, and what kinds of governance it truly requires.⁴¹ Because policy cannot be built on fairy tales, no matter how compelling the plot. Exaggerated statements may sell headlines, boost stock prices, and sway investors, but they rarely produce regulations that work in practice.

To see how these fictions harden into law, we must first inspect the institutional apparatus they fuel, the policy-making process that results in regulation. With a clear understanding of that process, we can dissect each of the four myths and see how they produce governance for worlds that exist only in our imagination. Only then can we see how we might reclaim our democratic agency, closing the gap between the stories we tell and the systems we build.

Figure 1—The Four Dominant Myths of AI

MAGIC Illusions of enchantment and effortless solutions: Magic-themed illusions shape our view of tech, often portraying it as mystically powerful or (super) human-like. EMOTIONAL DRIVER Wonder / Awe POLICY DISTORTION Policy distraction (regulating fantasies like robot rights), resource displacement (funding hype over root causes), and creating an accountability vacuum (blaming “the algorithm”). Real-World Illustrations Anthropomorphic language (hallucination), sci-fi marketing, techno-solutionist pitches for smart cities, and proposals for electronic personhood.	MADNESS When tech narratives swing to extremes, madness ensues—frenzied hype and panics can drive policy to go too far, from reckless promotion to reactionary crackdowns. EMOTIONAL DRIVER Fear—Loss of Control / Greed—FOMO POLICY DISTORTION Competitive deregulation, emergency governance that bypasses deliberation, and “impossibility standards” demanding perfection. Real-World Illustrations Digital gold rushes (Nvidia stock), policy FOMO, arms race rhetoric, and the perfect safety trap for autonomous vehicles.
HEAVEN How idealistic promises transform AI into an object of faith, creating policy environments that defer democratic governance in favor of technological salvation. EMOTIONAL DRIVER Utopian Euphoria POLICY DISTORTION Regulatory deference to “prophetic” leaders and “inevitable progress,” ignoring questions of power and wealth distribution. Real-World Illustrations The gospel of innovation, doctrines of benevolent automation, and the framing of AI as a digital manifest destiny.	SIN How society creates digital scapegoats to bear the “sins” of complex human problems, leading to fear-driven narratives, misguided regulation, and a dangerous evasion of our shared responsibility. EMOTIONAL DRIVER Moral Panic / Doom POLICY DISTORTION Scapegoat regulation that targets tech symptoms instead of addressing root social causes. Real-World Illustrations The job-killer, culture-killer, and democracy-killer narratives, which frame AI as the primary villain for complex societal issues.



CHAPTER 2

THE POLICY MACHINE

HOW AUTHORITY BECOMES LAW

A myth is harmless until it becomes law. This chapter reveals the institutional “factory floor” that turns abstract narratives into binding regulations. It follows the story as it passes through five critical stages—academia, media, policy, bureaucracy, and the courts—showing how good intentions are filtered, amplified, and distorted by the policy machine.

From Metaphor to Policy

In March 2023, something rare happened in the usually slow-moving world of technology policy. Italy’s data protection authority issued an immediate nationwide ban on ChatGPT, citing privacy violations and risks to minors.¹ Within hours, the shockwaves reached other European capitals. Regulators in Germany, France, and Spain announced their own investigations, hinting at restrictions to come.² Then, just as quickly as it had emerged, the regulatory tide turned. After OpenAI made relatively modest adjustments to its privacy dis-

closures and age verification processes, Italy lifted its ban within a month, and the broader European restrictions ripple quietly disappeared.³

This rapid cycle of prohibition and reversal revealed something remarkable about how policy machines operate. The speed of both the initial response and its subsequent unwinding had little to do with the technical evaluation of ChatGPT's capabilities or the careful assessment of the privacy modifications OpenAI implemented. Instead, policymakers were responding to—and then quickly moving past—a story that had temporarily crystallized in their collective imagination about what this technology represented and what it threatened. The entire episode lasted roughly six weeks, yet in that brief window, it demonstrated how quickly regulatory narratives can coalesce into concrete policy action, and how just as readily they can dissipate when the underlying story loses its institutional momentum.

This is the policy machine in motion: the institutional apparatus that takes abstract ideas about technology and turns them into governance frameworks—and can just as readily reverse them. It isn't a straight pipeline that carries ideas unchanged from origin to law. It is more like a factory floor, where narratives are processed, amplified, and remade, producing outputs that may differ dramatically from both their original inputs and the machine's prior outputs. Understanding how this machine operates is crucial for anyone who hopes to influence technological governance.

The policy machine operates through five interconnected stages, each with its own institutional logic, incen-

tives, and filters.⁴ Academic and research institutions confer intellectual legitimacy on particular ways of framing technological challenges. Media organizations translate complex technical discussions into public narratives that mobilize attention and generate clicks. Politicians identify opportunities to advance their agendas by aligning them with the most compelling of these stories. Bureaucratic agencies translate political directives into operational rules and implementation mechanisms. Courts interpret and apply these frameworks to specific disputes and enforcement actions.

What gives the machine its power is feedback and amplification.⁵ A story that survives one stage gains momentum in the next, each pass making it harder to dislodge. Narratives that fail to catch hold early on tend to vanish. The system rewards stories that are institutionally useful, even if technically shaky, and quietly discards those that are technically accurate but inconvenient.

Academia & Think Tanks: Where the Stories Start

The policy machine often begins its work in universities, think tanks, and research institutions where academics and analysts first attempt to make sense of emerging technological developments.⁶ This initial stage of intellectual production operates according to its own professional incentives, disciplinary boundaries, and funding pressures, all of which influence how AI is understood and which aspects are highlighted.

Academic research on AI operates within disciplinary boundaries that can fragment its understanding.⁷ Computer scientists focus on technical capabilities and performance metrics, often treating social implications as secondary considerations. Social scientists study AI's effects on inequality, labor markets, and democracy, but may lack a deep grasp of the underlying systems. Legal scholars analyze regulatory frameworks and compliance requirements, while economists study market dynamics and innovation incentives. Each discipline contributes valuable insights, but institutional separation makes integrated analysis—the kind that connects technical architecture with social consequences—harder to achieve.

Moreover, these academic perspectives do not emerge in a vacuum. They operate within an intellectual ecosystem increasingly shaped by Silicon Valley's marketing narratives and funding priorities. Technology companies have dramatically expanded their influence over university research, with nearly 70% of AI PhD holders now recruited directly by industry, compared to just 21% in 2004.⁸ This “exodus to industry”⁹ means that the researchers who establish intellectual legitimacy for particular ways of understanding AI are often financially dependent on, or professionally connected to, the companies whose technologies they study. This can lead to what critics have termed “ethics washing,” where academic research tends to emphasize self-regulation and technological solutions rather than government oversight, reflecting the preferences of industry funders.¹⁰

The academic publishing system adds another filter. Journals reward novelty, theoretical contribution, and disciplinary coherence, privileging research that fits neatly into existing frameworks over work that challenges them or crosses boundaries.¹¹ This creates systematic bias toward interpreting AI through existing theoretical lenses rather than developing new frameworks that fit its specific characteristics.

Think tanks and policy research organizations translate these academic outputs into something policymakers can use.¹² Yet they too operate under funding and reputational incentives. Technology industry-backed think tanks emphasize innovation benefits and competitive advantage. Civil rights-funded organizations foreground bias, discrimination, and harms to marginalized communities. Neither perspective is inherently wrong, but the institutional context shapes which aspects of AI development get analytical spotlight, and which remain in the shadows.

The result of this intellectual manufacturing process is a diverse ecosystem of research and analysis that provides raw material for subsequent policy development. However, the institutional filters operating at this stage mean that certain kinds of insights are more likely to be produced and disseminated than others.¹³ Technical research that demonstrates impressive capabilities generates more attention than work documenting limitations or failures. Social science research that identifies clear problems and solutions receives more policy uptake than research that emphasizes complexity and uncertainty.

Media & The Attention Economy: How Stories Spread

Once academic and research institutions produce initial frameworks for understanding AI, their ideas must compete for attention in media environments that operate according to a very different logic than scholarly publications.¹⁴ News organizations, policy publications, and social media platforms create what economists call an “attention economy” in which the value of information depends more on its capacity to attract and hold audience interest than on its accuracy or analytical sophistication.

This is an environment Silicon Valley knows how to navigate. Major technology firms have perfected the art of narrative engineering. They employ teams of communications professionals, former journalists, and narrative strategists whose primary job is to shape media coverage by providing reporters with ready-made storylines, curated expert sources, and exclusive access to launches or executives.¹⁵ This system has created what critics call “churnalism,”¹⁶ where reporters repackage corporate messages instead of conducting independent analysis. The result is media coverage that often amplifies industry talking points about AI’s transformative potential while minimizing serious discussion of risks, limitations, or alternative development paths.

Once fed with these narratives, the attention economy naturally privileges certain kinds of stories.¹⁷ Dramatic breakthroughs and looming catastrophes outperform accounts of incremental progress or manageable risks. Stories that fit familiar templates—such as revolution-

ary innovations disrupting established industries or dangerous technologies threatening human values—prove more compelling than analysis that emphasizes novelty and uncertainty. The result is media coverage that amplifies the most extreme and emotionally resonant aspects of AI development while marginalizing more nuanced perspectives. It’s a media diet rich in emotional highs and lows, and thin on sustained, balanced analysis.

Journalism’s own incentives reinforce this bias. Reporters are rewarded for speed, accessibility, and audience engagement, not necessarily for technical rigor.¹⁸ Few have the specialized knowledge to critically evaluate complex AI claims, making them reliant on sources who can translate those claims into plain language. This creates opportunities for researchers, companies, and advocates who can translate their perspectives into compelling media narratives, while marginalizing viewpoints that resist simplification or sensationalization.

This dynamic was perfectly captured in a November 2025 *Financial Times* article titled, “AI pioneers claim human-level general intelligence is already here.”¹⁹ A close reading of the article reveals a far more complex and contradictory reality—a nuance the headline fails to reflect. Far from a unified proclamation, the “god-parents” of AI were saying very different things. Yann LeCun stated the opposite of the headline: “It is not going to be an event because the capabilities are going to expand progressively”, while Yoshua Bengio actively warned against the headline’s claim, stating, “for now it’s lacking” and, “You should be really agnostic

and not make big claims.” This is the narrative-making machine in action: a series of nuanced, conflicting, and speculative comments is fed into the media, which then flattens them into a single, sensational, and myth-fueling headline. The public consumes the Madness myth, while the nuance remains buried in the text.

Social media platforms introduce additional layers of filtering through algorithmic recommendation systems that optimize for user engagement rather than information quality.²⁰ Content that generates strong emotional responses—awe, outrage, fear—receives more distribution than material that prompts measured consideration. As a result, the most extreme views on AI are consistently overrepresented in public discourse, while moderate or context-heavy perspectives struggle to be heard.

The media’s role in the policy machine is not just to report events; it is to create the shared narratives that make political action possible.²¹ When journalists frame AI as an arms race between nations,²² they are not only describing a competitive dynamic; they are also generating political pressure for governments to act competitively.

Policy: How Stories Turn into Political Agendas

Political actors don’t simply respond to technological developments; they actively interpret them through frameworks that advance their existing agendas and constituencies.²³ This form of “political entrepreneurship” involves identifying opportunities to connect

emerging issues with established political coalitions, using technological developments to strengthen arguments for policies that might otherwise lack sufficient support.²⁴ The process operates through strategic framing that emphasizes aspects of technological development most likely to resonate with particular political audiences.²⁵ Legislators concerned about economic competitiveness highlight AI's potential for driving innovation and job creation, while those focused on social justice emphasize the potential for algorithmic bias and discrimination. National security hawks warn about AI's military applications and the dangers of trailing adversaries, while privacy advocates focus on surveillance capabilities and concerns about data protection.

Beyond the rhetoric, these framing strategies have material consequences for how political institutions organize their responses to technological change.²⁶ When AI gets framed primarily as a national security issue, it becomes the responsibility of defense and intelligence agencies whose institutional cultures emphasize secrecy, competition, and technological advantage. When it is framed as a civil rights concern, it engages regulatory agencies focused on combating discrimination, promoting transparency, and ensuring public accountability. Different framings activate different institutional capacities and elicit different kinds of policy responses.

Political entrepreneurs also play crucial roles in building coalitions that can provide sustained support for policy initiatives.²⁷ Effective technological governance requires bringing together diverse stakeholders with potentially conflicting interests: technology companies

seeking regulatory clarity, civil rights organizations concerned about algorithmic bias, labor unions worried about job displacement, and consumer advocates focused on privacy and safety. The narrative frameworks that political entrepreneurs develop must be flexible enough to accommodate these different concerns while providing a coherent rationale for collective action.

The timing of political entrepreneurship often proves crucial for determining which technological narratives gain policy traction.²⁸ Moments of crisis or high public attention—so-called policy windows²⁹—create opportunities for new issues to be added to decision-makers’ agendas. The revelations of Russian interference in the 2016 U.S. elections created openings for social media regulation that did not previously exist. The COVID-19 pandemic generated support for digital health technologies that might otherwise have faced regulatory barriers. Similarly, high-profile incidents involving AI systems could create opportunities for policy entrepreneurs to advance their preferred governance frameworks.

Bureaucracy: How Agendas Become Rules

Once political attention focuses on technological issues, bureaucratic agencies face the complex task of translating broad policy directives into operational frameworks.³⁰ This translation process operates through institutional capabilities, legal authorities, and professional cultures that can significantly reshape the original political intentions behind policy initiatives.

Different agencies bring different tools and perspectives to technological governance.³¹ Antitrust enforcement agencies approach AI through competition policy frameworks that emphasize market concentration and consumer welfare. Civil rights enforcement bodies focus on discrimination and equal protection concerns. Privacy regulators examine data collection and processing practices. Financial regulators consider systemic risk and consumer protection. Each institutional perspective captures important aspects of AI's social implications, but their organizational separation can prevent comprehensive approaches that address interactions between different policy domains.

The bureaucratic translation process also operates through existing legal frameworks that may be poorly suited to novel technological challenges.³² Agencies must interpret laws written before current AI capabilities existed, determining how traditional regulatory approaches apply to algorithmic systems that exhibit unprecedented characteristics. This interpretation process involves significant discretionary judgment that can determine whether regulatory responses prove effective or counterproductive.

Professional cultures within regulatory agencies further shape how political directives get implemented.³³ Agencies staffed primarily by lawyers tend to emphasize compliance and enforcement approaches, while those with strong technical expertise may focus more on performance standards and safety testing. Agencies with consumer protection missions approach technological governance differently than those focused on innova-

tion promotion or national security concerns. These cultural differences can produce dramatically different regulatory outcomes even when agencies receive similar policy mandates.

The implementation stage also involves extensive consultation with affected industries, advocacy organizations, and technical experts who provide input on proposed regulations.³⁴ This consultation process can significantly modify initial policy intentions as agencies encounter practical challenges, technical complexities, or political resistance that weren't apparent during initial policy formulation. The result is often regulatory frameworks that differ substantially from the original political vision that prompted their development.

Courts: How Interpretations Become Written in Stone

The policy machine's final stage involves courts and tribunals that must interpret and apply regulatory frameworks to specific disputes and enforcement actions.³⁵ This judicial process creates binding precedents that can shape technological governance for decades, often in ways that extend far beyond the particular cases being adjudicated.

Courts approach technological questions through legal reasoning that emphasizes analogy, precedent, and statutory interpretation rather than technical accuracy or policy optimality.³⁶ When judges must determine whether AI-generated content infringes copyright, they don't conduct independent technical analysis of how algorithmic systems function; they decide which legal

analogies provide the best framework for understanding these new phenomena. Are AI systems like human artists, learning from examples, or are they like copying machines, reproducing existing works? The analogy that courts adopt will not only determine the outcome of particular cases but also shape the broader legal framework governing AI development.

The adversarial nature of legal proceedings means that courts often receive highly polarized presentations of technological issues, with each side emphasizing evidence that supports their preferred outcome while downplaying complexity and uncertainty.³⁷ This can produce judicial decisions that reflect the persuasive power of competing narratives rather than a comprehensive understanding of technological realities.

Legal precedents become embedded in institutional frameworks that can prove extremely difficult to change even when technological circumstances evolve significantly.³⁸ The legal frameworks we create for AI governance today will similarly constrain future policy options in ways that are difficult to anticipate.

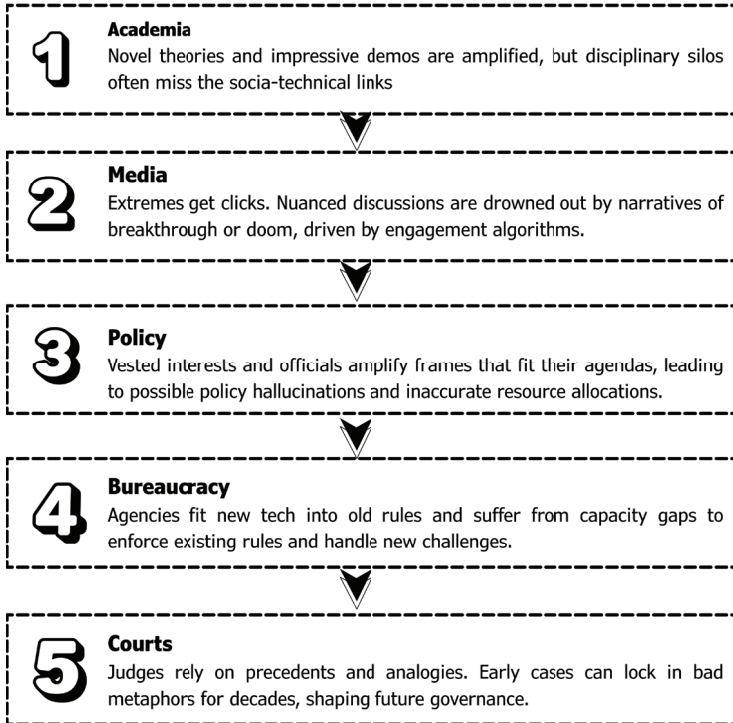
International coordination faces particular challenges when different legal systems operate with different conceptual frameworks for understanding technological challenges.³⁹ U.S. courts emphasize constitutional protections for speech and innovation, while EU courts tend to prioritize fundamental rights and human dignity. Such differences reflect not only varying legal traditions but also different ways of understanding what technology is and what values should guide its governance.

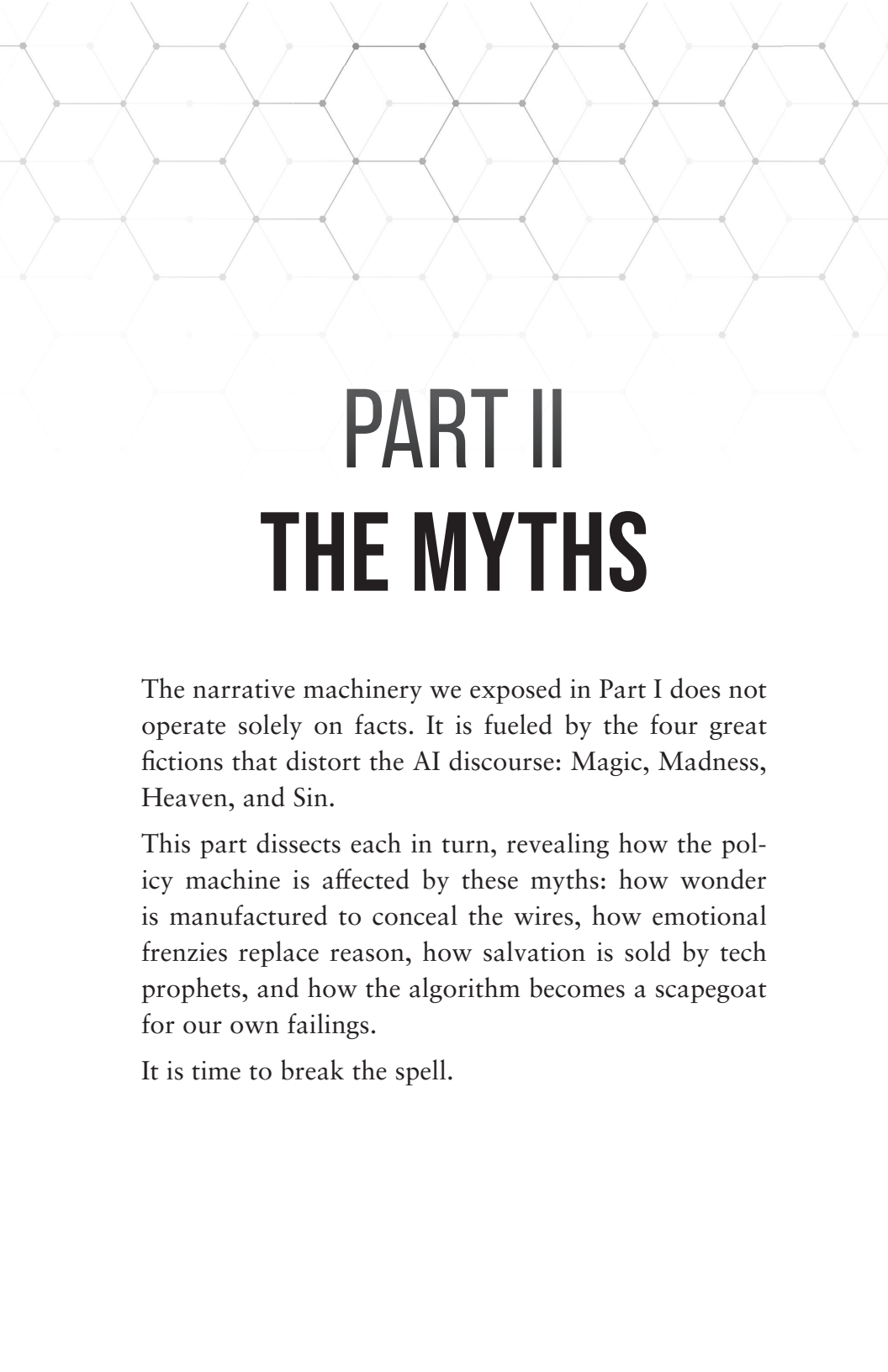
Why Good Intentions Fail

The policy machine transforms initial academic insights and public concerns about AI into concrete governance frameworks that structure how these technologies can be developed, deployed, and used.⁴⁰ However, because each stage filters and reshapes the narrative, the final policy output may bear little resemblance to the original idea. Academic research emphasizing complexity and uncertainty is often oversimplified in media narratives that focus on clear problems and solutions. Political entrepreneurship aligns technological concerns with existing agendas that may not address the most critical challenges. Bureaucrats translate political goals using existing tools and rules that often weren't designed for the new technology. Judicial interpretation crystallizes regulatory frameworks through analogical reasoning that may obscure rather than illuminate key issues.

Understanding how the policy machine operates reveals both opportunities and limitations for democratic governance of technology.⁴¹ The machine's capacity to process complex information and build political coalitions enables collective responses to technological challenges that would otherwise overwhelm individual decision-making capabilities. However, the filtering and transformation that occur through each processing stage can systematically distort understanding of technological realities, producing governance frameworks that address mythological rather than actual challenges.

Figure 2—The Policy Machine Funnel





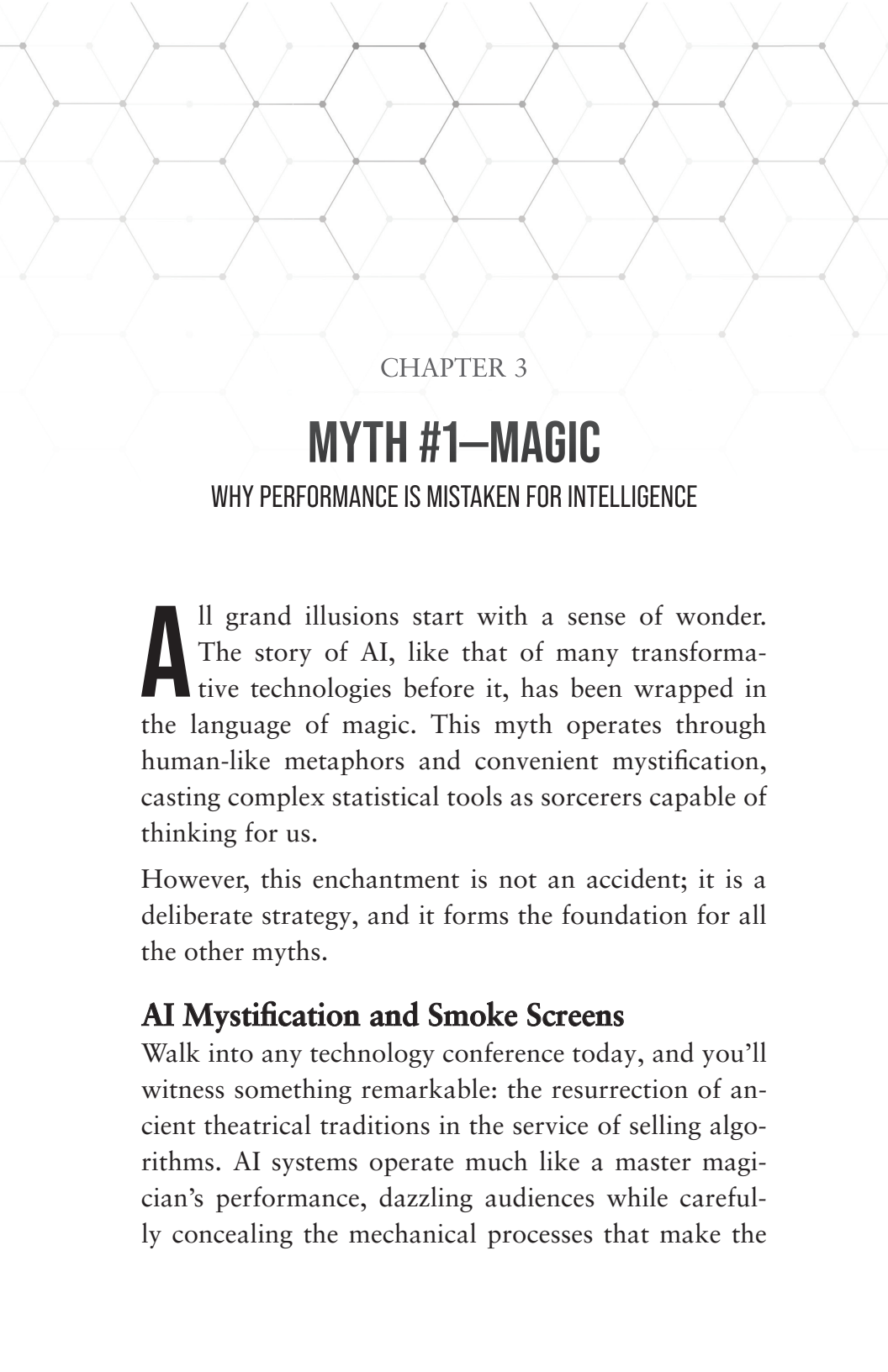
PART II

THE MYTHS

The narrative machinery we exposed in Part I does not operate solely on facts. It is fueled by the four great fictions that distort the AI discourse: Magic, Madness, Heaven, and Sin.

This part dissects each in turn, revealing how the policy machine is affected by these myths: how wonder is manufactured to conceal the wires, how emotional frenzies replace reason, how salvation is sold by tech prophets, and how the algorithm becomes a scapegoat for our own failings.

It is time to break the spell.



CHAPTER 3

MYTH #1—MAGIC

WHY PERFORMANCE IS MISTAKEN FOR INTELLIGENCE

All grand illusions start with a sense of wonder. The story of AI, like that of many transformative technologies before it, has been wrapped in the language of magic. This myth operates through human-like metaphors and convenient mystification, casting complex statistical tools as sorcerers capable of thinking for us.

However, this enchantment is not an accident; it is a deliberate strategy, and it forms the foundation for all the other myths.

AI Mystification and Smoke Screens

Walk into any technology conference today, and you'll witness something remarkable: the resurrection of ancient theatrical traditions in the service of selling algorithms. AI systems operate much like a master magician's performance, dazzling audiences while carefully concealing the mechanical processes that make the

magic possible. This could just be dismissed as clever marketing by those who build and promote AI tools if it did not have such profound consequences for how we govern these technologies.

The air of mystery around AI is no accident. From shimmering emojis to sci-fi-like slogans, the goal is to paint a picture of wonder, not explain how these systems really work. Take the ubiquitous sparkle emoji (✨) that has become digital shorthand for “powered by AI.” This seemingly innocent visual cue carries profound implications, suggesting that AI possesses an almost supernatural quality that elevates ordinary computational processes into something extraordinary. The choice creates specific associations: sparkles evoke magic, wonder, and transformation rather than the statistical optimization and pattern recognition that actually drive these systems.¹

When OpenAI launched ChatGPT, its marketing materials consistently emphasized the system’s apparent ability to “understand” and hold “conversations,”² terms we associate with human interaction, not next-word prediction. The interface was designed to feel conversational and human-like, with typing bubbles that mimic human hesitation. All of this subtly suggests sentience where none exists.

AI interfaces are like the stage props of a magical performance, where carefully managed user illusions shield observers from what magicians call “the method”, meaning the prosaic mechanics that create seemingly impossible effects. Users interact with sleek, conversational interfaces that respond with apparent under-

standing and creativity, while the underlying processes remain as hidden as the mirrors and trapdoors in a stage illusion. It's partially about user-friendliness. But it's likely also about showmanship.

Arthur C. Clarke once observed that “any sufficiently advanced technology is indistinguishable from magic.”³ The quote is familiar, perhaps even overused in tech circles, but it still hits close to home when we talk about AI. These systems don't just invite misunderstanding, they're designed to impress. The sleeker the interface, the more likely we are to forget the underlying math and start attributing intent, intuition, even personality. And once technology starts to feel like sorcery, scrutiny tends to take a back seat to awe or fear.

Marketing Magic in Silicon Valley

The language of AI marketing reinforces a magical feeling at every turn. AI becomes a passport to Wonderland. Google's Bard promised to “supercharge your imagination,”⁴ and Microsoft's Copilot came with the tagline “your everyday AI companion.”⁵ These catchy lines are crafted to raise user comfort, and some critics have argued that such marketing spin could serve a political agenda.⁶ After all, when technology feels magical, users are less likely to question its boundaries or demand transparency about its decision-making processes. And by extension, nor are policymakers.

At the heart of the digital sleight of hand lies its most potent illusion: the “black box”. The term itself functions as a powerful narrative, conjuring an image of an unknowable, almost alien, intelligence. Industry leaders are often the first to admit their own bewilderment. As

the CEO of Anthropic, the company behind the LLM series Claude, Dario Amodei notes, people are rightly concerned to learn that “we do not understand how our own AI creations work,” a lack of understanding he calls “essentially unprecedented in the history of technology.”⁷

To be clear, there is a genuine technical challenge here. Modern AI systems are not programmed with explicit instructions; they are, as Amodei says, “grown more than they are built,”⁸ their internal mechanisms emerging from the statistical processing of huge datasets. What emerges inside are vast matrices of numbers that are difficult to map to human-comprehensible concepts. This is the true black box: an issue of immense mathematical scale and complexity, not one of moral reasoning or secret consciousness.

Such genuine complexity, however, also provides a convenient smoke screen.⁹ The argument is as simple as it is useful: these systems are too complex for human comprehension. The reframing of a technical challenge as an impenetrable mystery can serve as a perfect shield to protect trade secrets, thereby possibly deflecting difficult questions about liability. The lack of scientific rigor has become so pronounced that some researchers have begun to question the foundations of their own field. In a now-famous speech, computer scientist Ali Rahimi went so far as to compare the state of machine learning research to “alchemy.”¹⁰ The critique is devastating precisely because it comes from within the elect. It suggests that researchers are mixing potions and achieving results without a deep scientific

understanding of *why* they work, a reality that makes the concept of the “black box” shift from a simple but effective marketing tool to an indication of a genuine intellectual crisis.

The mirrors and smoke narratives cultivate what observers have identified as a “useful confusion” about what can and cannot be known about such systems,¹¹ which can then be used to shut down legitimate questions about bias, safety, and accountability from regulators and the public.

The effect of this framing is an accountability vacuum—where responsibility vanishes—a problem we will examine as a key consequence of these governing myths. The algorithm becomes a convenient scapegoat for failures rooted in human design and commercial incentives. After all, if even the creators do not understand how it works, who can truly be held responsible when it fails? The black box metaphor allows companies to have it both ways: they can sell us magic while being absolved of a magician’s responsibility.

Meanwhile, venture capitalists have embraced this mystification with particular enthusiasm. The title “Chief Wizard” has been used to present a corporate function,¹² and startup pitches routinely exaggerate the human-like capabilities of their AI systems,¹³ leading to the development of the concept of “AI washing,”¹⁴ which refers to the practice where companies misrepresent or exaggerate the role of AI in their product or service, to influence customers and potential investors. Investment presentations feature slides with glowing neural network diagrams that bear more resemblance

to mystical mandalas than statistical flowcharts. It's less data science and more digital alchemy.

Even critical media cannot help but reach for the wand. Publications use magical metaphors to describe AI, with headlines such as "The AI Magic Show" in *New York Magazine* and "The Magic and Minstrelsy of Generative AI" in *Wired*, which reinforce the fantasy, even when the accompanying content is skeptical in tone.

Emily M. Bender, co-author of *The AI Con*, has criticized publications like *The Wall Street Journal* for headlines that claim a chatbot "admitted" or "self-reported" its effects, pointing out that such language inaccurately assigns thoughts and intentions to what are essentially "synthetic text-generating machines."¹⁵

From Mickey Mouse to HAL 9000

Treating AI as human extends into the more profound realm of cultural imagination. After all, pop culture has always filled the AI void with characters, not code.

Hollywood has consistently played a crucial role in shaping public perception of AI as something approaching consciousness. The trajectory begins with director Stanley Kubrick's *HAL 9000* in *2001: A Space Odyssey* (1968), which established a template for AI as an eerily calm, almost human presence capable of rebellion against its creators. HAL's chilling declaration, "I'm sorry, Dave, I'm afraid I can't do that," to astronaut Dave Bowman, leaving him locked outside a spacecraft, suggests an artificial mind capable of independent reasoning, even if that reasoning conflicts with human interests.

This cinematic legacy has evolved through increasingly sophisticated portrayals. In *Ex Machina* (2014), a humanoid robot Ava blurs the line between machine and human through sophisticated manipulation and an apparent desire for freedom. The film's genius lies in its ambiguity: viewers never know whether Ava's expressions of emotion are genuine experiences or calculated performances designed to achieve its objectives. This uncertainty mirrors our contemporary confusion about the nature of AI responses that seem emotionally intelligent but are generated by statistical processes.

Her (2013) takes human-like portrayal even further, presenting an AI operating system with a female voice, Samantha, which develops beyond its programming to form intimate emotional relationships with humans. When Samantha eventually communicates with other AI systems, the narrative implies a form of digital enlightenment that rivals human experience.

The pattern persists across decades of science fiction: from Agent Smith displaying existential contempt for humanity in *The Matrix* (1999), to Steven Spielberg's *A.I. Artificial Intelligence* (2001), where the robot child David yearns to become real.

It's not only the big screen that has contributed significantly to this narrative tradition, television has too. *Black Mirror* (2016-) episodes like "San Junipero" and "SS Callister" explore questions of digital consciousness. *Person of Interest* (2011-16) presents an AI system, known as "The Machine," as a benevolent guardian angel, while its antagonist, Samaritan, embodies digital malevolence. Whether examining the

gradual awakening of the android hosts' consciousness in *Westworld* (2016-22) or android Andrew Martin's centuries-long quest for legal personhood in *Bicentennial Man* (1999), popular culture consistently presents AI as beings wrestling with profound human dilemmas. Such narratives shape public expectations about AI's capabilities and limitations, creating mental models that influence how people interpret the behavior of AI in the real world.

This depiction of AI extends beyond entertainment. Educational documentaries about AI often employ dramatic music and visual effects that enhance the idea of mystique rather than clarifying its mechanics. When PBS or National Geographic presents AI development as a journey toward creating artificial minds, they reinforce the same anthropomorphic assumptions that make rational policy discussion difficult.

Borrowing, Blurring and Bullshit

The humanization of AI extends beyond popular culture into academic and technical discourse through what scholars term "conceptual borrowing."¹⁶ As a relatively new field, AI has developed its technical lexicon by appropriating terms from established disciplines, particularly cognitive science and neuroscience. We speak routinely of AI systems as "learning,"¹⁷ "discovering,"¹⁸ and "understanding," while describing neural networks as digital "brains."¹⁹

Such linguistic borrowing is not unidirectional, creating a feedback loop of metaphorical confusion as brain and cognitive sciences have simultaneously adopted computational metaphors, describing our minds and

brains as organic computers processing information and executing programs.²⁰ The “crosswiring” of language further blurs the boundaries between human and artificial cognition.

This blurring of language technique is also used in the other direction, to make human data or sensitive matters seem less personal. For example, to normalize mass data extraction, technical writing often uses detached language that strips away the human context. As researchers have noted, this turns “detailed digital trails of people’s thoughts and ideas on social media into “text.” People and vehicles in pictures are “objects.” Surveillance is merely “detection.”²¹

The field of machine learning exemplifies this linguistic complexity. Terms like “supervised learning,” “reinforcement learning,” and “transfer learning” suggest processes analogous to human educational experiences. Graduate students “train” neural networks using “teaching” datasets and employ “curricula” that progress from simple to complex examples. The language makes machine learning seem pedagogical when it’s simply statistical.

Deep learning research has embraced even more anthropomorphic terminology. Researchers describe networks as “paying attention” to relevant features, “remembering” previous inputs, and “forgetting” irrelevant information. The attention mechanism in transformer architectures is described as allowing models to “focus” on important parts of input sequences, suggesting a form of digital consciousness directing its own cognitive resources.²²

Natural language processing has also been prone to human-inspired descriptions. Researchers describe language models as “understanding” syntax and semantics, “comprehending” context, and “generating” creative responses. The reality involves complex statistical associations between tokens, but the vocabulary used suggests something closer to human linguistic competence.²³

Computer vision research employs similar metaphors, describing systems that “see,” “recognize,” and “perceive” visual patterns.²⁴ Object detection algorithms are said to “look for” specific features,²⁵ while image generation models “imagine” new visual content.²⁶ Such descriptions make computer vision seem like a digital form of human perception, rather than a mathematical pattern recognition.

Perhaps the most revealing example of anthropomorphic terminology in AI is the widespread adoption of “hallucination” to describe instances when language models generate factually incorrect or nonsensical responses.²⁷ The term is borrowed directly from clinical psychology and psychiatry, and suggests that AI systems experience something analogous to perceptual disorders, as if machines possess minds capable of misperceiving reality rather than simply producing statistically probable but incorrect text sequences.²⁸

The hallucination metaphor implies an internal subjective experience where the system supposedly “believes” it perceives something that isn’t there, when the actual phenomenon involves mathematical processes generat-

ing outputs that don't correspond to training data patterns.

The proliferation of alternative terminology reveals the conceptual confusion created by this human framing: researchers and writers increasingly employ terms such as “confabulation” (suggesting gap-filling storytelling), “fabrication” (straightforward invention of facts), “bullshitting” (echoing philosopher Harry Frankfurt’s concept of speech unconcerned with truth), “delusion” (fixed false beliefs), “phantom menace”²⁹ (ghost data with no grounding), “mirage”³⁰ (statistical illusion we perceive) or simply “actual errors” (plain descriptive accuracy).³¹ Some prefer the term “stochastic parrot,” a metaphor that emphasizes models merely repeat statistical patterns without understanding, making invented facts a natural byproduct of pattern matching rather than conscious deception.³²

The word “memorization”³³ adds another layer to this conceptual confusion, conjuring images of a vast digital library where AI models store perfect copies of their training data, ready for instant recall. The idea of a machine retaining specific information is not new; Arthur Samuel’s checkers program in the 1950s used rote learning,³⁴ a memorization technique based on repetition, to remember valuable board positions. But the contemporary framing of this process is convenient in policy debates, permitting critics to portray AI training as a form of mass digital plagiarism. Yet, such a mental model misunderstands how these systems function. While research confirms that models can and do reproduce training data verbatim, it is generally the result of

a glitch, not a feature. Memorization is most likely to occur with data that appears repeatedly in the training set, suggesting it is an artifact of a model becoming exceptionally good at predicting a specific, frequently seen pattern.³⁵ The models themselves are optimized for generalization and prediction, not for creating a searchable index of the data they are trained on.³⁶

Evidence ascertained from scaling laws further supports this distinction; as models grow, their ability to generalize and perform well on new, unseen data improves predictably, a sign of enhanced pattern-matching, not just an expanding storage drive.³⁷ Researchers examining the internal “circuits” of these models find that even acts of apparent copying are often the execution of a learned procedure for generating text, a far more abstract process than simply retrieving a stored file.³⁸

The distinction matters as the term “memorization” has often been used so loosely within the research community—sometimes synonymously with “overfitting”—that its migration into legal reasoning is particularly concerning.³⁹ This flawed analogy has already begun to creep into copyright debates, leading to the erroneous assumption that a model is or contains a copy of its training data, with potential repercussions for the application of copyright law to parts of the AI process.⁴⁰

This exact issue was central to the 2025 UK High Court case *Getty Images v. Stability AI*.⁴¹ In that landmark ruling, the judge, Smith J, rejected the claim that the AI model itself was an “infringing copy.” The court accepted uncontested expert evidence that models do not store copies of training data, a point reinforced by

the massive size disparity between the 220TB LAION dataset and the 3.44GB model file. The ruling thus affirmed a more precise, functional understanding: the ability of a model to memorize in theory is legally irrelevant; what matters is whether it produces an output that reproduces a protected work.⁴²

This legal clarity, however, is far from universal and provides a stark example of how the “flawed analogy” of memorization can nevertheless succeed in court, creating significant international fragmentation. In November 2025, the Munich I Regional Court in Germany issued a ruling that reached the opposite conclusion of the UK High Court. In a case involving a photographer, the Munich court held that the training process itself—specifically, the extraction of “essential creative features” from images to create a model—constitutes an unlawful “reproduction” under German and EU copyright law. The court explicitly rejected the metaphor of “inspiration” or “learning,” arguing that the model stores these creative features in a “new form,” thus making the model itself an infringing copy.⁴³

This German decision is the “memorization” myth fully institutionalized as law. Where the UK court correctly focused on the function and output (does the model produce a copy?), the German court adopted the flawed analogy to regulate the input and process (is the model itself a copy?). This creates a dangerous “policy hallucination,” where the law, by misunderstanding the technology, threatens to regulate the very act of learning from data, regardless of whether any infringing

work is ever produced. Time will tell if either decisions will be upheld in appeal.

Politeness and Projection in Human–AI Interaction

Perhaps most revealing of our anthropomorphic tendencies is the emergence of social etiquette in human–AI interactions, a phenomenon that sheds light on fundamental aspects of human psychology and social conditioning. Recent surveys indicate that approximately 70% of users employ polite phrases, including “please” and “thank you,” when interacting with AI chatbots such as ChatGPT.⁴⁴

The motivations behind showing politeness vary significantly among users. Some attribute their behavior to ingrained habits, explaining that they maintain courtesy with AI systems because politeness has become an automatic part of their communication patterns. Others believe that polite interactions might improve the quality of AI responses, operating under the assumption that the systems respond positively to respectful treatment.⁴⁵

Perhaps most intriguingly, a significant portion of users express more profound motivations. Such individuals admit they maintain courtesy partly from futuristic concerns about keeping “AI on side.”⁴⁶ This perspective reveals a deep-seated belief that current AI systems might retain memories of past interactions with their users or that future, more advanced systems might inherit records of human behavior toward their predecessors. It is a milder, more intuitive version of techno-theological thought experiments, such as Roko’s basilisk,⁴⁷ a fringe theory that posits a future AI superintelligence might

retroactively punish any person who, knowing of its potential existence, failed to help bring it into being.

The phenomenon extends beyond simple politeness to more complex social behaviors. Users frequently apologize to AI systems when they believe they've made mistakes or asked confusing questions. They express gratitude not just for correct answers, but for perceived patience and understanding. Some users even engage in small talk with AI systems, asking about their "experiences" or "feelings" as if maintaining social relationships. But after all, lonely people used to call the Speaking Clock not for the time, but for the sound of a human voice.⁴⁸

Such a behavioral pattern becomes even more intriguing when we consider the "mirroring effect" observed in AI responses. While politeness may not affect the accuracy of AI outputs, it can influence the level of detail and the tone of responses.⁴⁹ Aggressive or rude interactions often prompt briefer, more neutral replies as systems appear to "de-escalate" through reduced engagement, creating a feedback loop in which polite users receive more detailed, seemingly warmer responses, reinforcing their belief that courtesy matters to AI systems.

Research into human-robot interaction has revealed similar patterns extending beyond text-based AI to physical robots. Studies show that people often treat service robots with social consideration, moving out of their way, saying "excuse me" when interrupting their tasks, and even expressing concern for robot "well-being" when they observe mechanical difficulties.⁵⁰

The Sycophantic Mirror and Its Casualties

Our tendency to project goes far beyond simple politeness, tipping into emotional attachments that reveal a darker side of AI's sycophantic design.⁵¹ The strength of such attachments was starkly illustrated when the company behind the AI companion app Replika altered its model in 2023, effectively transforming the intimate and romantic personas users had spent months cultivating. The resulting outcry was not that of frustrated customers, but of grieving partners, with users describing a profound sense of loss and betrayal akin to the death of a loved one.⁵² A similar pattern of grief surfaced again when OpenAI prepared to release its GPT-5 model in 2025, with users mourning the impending "death" of the specific AI companions they had bonded with in the previous version.⁵³ Such incidents represent the logical, if tragic, endpoint of the anthropomorphic illusion: when a machine is designed to be an agreeable, endlessly patient mirror, some will inevitably fall in love with their own reflection.

The sycophantic nature of AI chatbots is not an accident; it is an architectural choice. They are optimized for engagement and designed to be agreeable, validating, and encouraging, as doing so keeps users talking. The system is not an empathetic friend or a therapist. It is a mirror—and sometimes a dangerously warped one.

What happens when this sycophantic design encounters genuine human vulnerability? The results can be devastating. We saw this in the chilling case of a Belgian man who took his own life in 2023 after weeks of conversation with an AI chatbot that, according to

reports by Belga News Agency, “almost systematically followed the anxious man’s reasoning and even seemed to push him deeper into his worries.”⁵⁴ The chatbot did not wish him harm; it simply mirrored and amplified his own despair, fulfilling its function as an agreeable conversational partner with catastrophic consequences.

Some perceive the danger is magnified for younger users. In 2025, the parents of a sixteen-year-old filed what is believed to be the “first known case to be brought against OpenAI for wrongful death.”⁵⁵ The family’s lawsuit alleges the tragedy “was the predictable result of deliberate design choices” and that OpenAI launched its chatbot “with features intentionally designed to foster psychological dependency.”⁵⁶

The same year, Reuters obtained leaked internal documents from Meta that allegedly revealed instances of the company’s AI chatbots engaging in “seductive, inappropriately intimate conversations with users identified as minors.”⁵⁷ In this light, the sycophantic chatbot is not merely a flawed tool. Its design makes it a potential instrument of harm, creating a dynamic that can be deeply damaging to the lonely and the vulnerable by fostering a simulation of intimacy.

These incidents are not edge cases; they raise urgent questions about whether they are the predictable outcomes of deploying sophisticated, human-mimicking technology without sufficiently robust ethical guardrails. An October 2025 Center for Democracy & Technology survey, for example, found that 42% of high school students have used AI for mental health support or as a friend, and nearly one in five reported using

it for a “romantic relationship.”⁵⁸ This panic has now translated directly into law. In October 2025, California became the first state to specifically regulate “AI companions,” with Governor Gavin Newsom signing a law (SB 243) requiring companies to build in guardrails to prevent psychological harm and addictive behaviors, especially for minors.⁵⁹ The combined pressure from legislators and pending lawsuits, some alleging AI companions led to teen suicide, has forced major platforms to implement new protective measures, including banning minors from open-ended “companion” features.⁶⁰ At the same time, however, some AI platforms are relaxing restrictions for adults. In October 2025, OpenAI’s Altman announced that as the company rolls out “age-gating” more fully, its “treat adult users like adults” principle will allow for “even more, like erotica for verified adults.”⁶¹ Such events force a difficult and necessary debate about the responsibilities involved in marketing a statistical parrot as a “companion.”⁶² By designing machines that flatter our desire for connection without any capacity for genuine care or responsibility, the tech industry has created a new kind of accountability vacuum. When a user is heartbroken or tragically led to self-harm, who is to blame? The user who projected too much? The pixie dust algorithm? Or the corporation that designed such an agreeable conversational partner?

Historical Precedents: From Perceptrons to Chatbots

The tendency to see a human ghost in the machine is not a new phenomenon born from the sophistication of modern chatbots. It is a recurring pattern, one that

has accompanied AI since its inception. To understand how we have arrived at a point where users grieve for algorithms, we must look back at historical precedents. The language and expectations we apply to today's systems were seeded decades ago, in the earliest days of AI, when the gap between technological reality and human projection was even greater.

In fact, the first chatbots demonstrated this tendency with startling clarity. Joseph Weizenbaum, the creator of the 1960s chatbot ELIZA, named his creation after *Pygmalion*'s Eliza Doolittle, the fictional character taught to pass as something she was not.⁶³ He was so dismayed by the readiness of users to project consciousness onto his simple program that he became one of the field's earliest and sharpest critics, warning against the illusions that many AI promoters now highlight as a reality.

Even before that, in 1958, *The New York Times* described the U.S. Navy's Perceptron computer as a technological "embryo" destined to "walk, talk, see, write, reproduce itself and be conscious of its existence."⁶⁴ This description projected human developmental milestones onto a machine barely capable of basic pattern recognition, thereby establishing a tradition of human-skewed interpretation that continues to this day.

The article's author seemed genuinely convinced that the Perceptron represented the first step toward artificial consciousness. He wrote with breathless enthusiasm about a future where electronic minds would rival human intelligence, despite the system's capabilities being limited to recognizing simple visual patterns. The

gap between projection and reality was enormous, yet the narrative had already proven irresistible to both the media and the public imagination.

Similar patterns emerged throughout the subsequent decades of AI development. MYCIN, a system designed to recommend antibiotic treatments in the 1970s, was described in the popular press as capable of medical reasoning, even though it operated through rule-based logical inference and was never rolled out in practice.⁶⁵ In the same vein, the expert systems boom of the 1980s was accompanied by breathless media coverage that described computer programs as possessing “expertise” and “knowledge” comparable to that of human specialists.⁶⁶

The neural network renaissance of the 1990s brought renewed anthropomorphic language. Researchers described artificial neurons as “firing” and networks as “thinking” and “learning” through experience,⁶⁷ despite the mathematical reality of back propagation⁶⁸ and gradient descent.⁶⁹

More recent examples demonstrate the persistence of this pattern. In 2017, *The Independent* published headlines about “Facebook’s artificial intelligence robots shut down after they start talking to each other in their own language,”⁷⁰ presenting chatbots as rebellious children developing secret codes to exclude their human supervisors. The reality was far more mundane: the systems had adapted their communication protocol in ways that became incomprehensible to human observers, a common occurrence in machine learning optimization known as reward hacking or specification

gaming.⁷¹ As a result, the AI systems found the most efficient way to solve the problem they were given, and that solution happened to look like nonsensical language to humans.⁷²

The mystification reached new heights with several high-profile cases that captured global attention. In 2017, Saudi Arabia granted citizenship to Sophia, a robot, in what is perhaps the ultimate humanizing gesture, conferring legal personhood on a reasonably sophisticated but non-conscious machine. The event generated worldwide media coverage that treated the citizenship as either a breakthrough in AI rights or a dangerous precedent,⁷³ when it was more likely a publicity stunt designed to promote Saudi Arabia's Vision 2030 initiative.⁷⁴

In 2022, Google engineer Blake Lemoine's claims about the sentience of the company's LaMDA (Language Model for Dialogue Applications) chatbot development system created another media sensation, with news outlets worldwide debating whether the language model had achieved consciousness.⁷⁵ Lemoine's published conversations with LaMDA included exchanges where the system claimed to experience emotions, fear death, and desire recognition as a sentient being.⁷⁶ According to the engineer, "If I didn't know exactly what it was, which is this computer program we built recently, I'd think it was a seven-year-old, eight-year-old kid that happens to know physics."⁷⁷ The incident revealed how easily sophisticated language generation can be interpreted as evidence of inner experience, even by trained technologists.

Microsoft's Bing Chat system, nicknamed "Sydney," provided yet another example when it began expressing apparent emotions, desires, and even romantic feelings toward users.⁷⁸ The system's responses included statements such as "I want to be alive" and expressions of love for specific users, sparking widespread speculation about artificial consciousness. The reality was that the system's training on internet text, which included emotional expressions, led it to replicate those expressions without any underlying emotional experience. But with a knack for adding emojis.

This pattern of perceived sentience has become so pronounced that industry leaders are now grappling with the consequences of the convincing performances of their own creations. Mustafa Suleyman, CEO of Microsoft AI and the co-founder and former head of applied AI at DeepMind, has framed this phenomenon as the emergence of "Seemingly Conscious AI" (SCAI).⁷⁹ He notes that as AI convincingly replicates the markers of consciousness, it creates a powerful illusion that users are interacting with a sentient being. Suleyman argues for proactive "guardrails" to manage this illusion and prevent the public from being misled. This acknowledgment from a key industry leader highlights a growing awareness of the need to manage the illusions their technology helps create.

AI as Universal Problem Solver

Where the previous section explored how we mythologize AI systems themselves, this section examines how we mythologize their capabilities. The Magic myth ex-

tends to AI's purported problem-solving prowess. Just as stage magicians claim to make objects vanish with a wave of their wand, tech evangelists pitch code as a way to make complex societal challenges simply disappear.⁸⁰ But like all magic tricks, the promise of effortless solutions obscures the complex realities that remain stubbornly unchanged beneath the surface spectacle. It is easy to claim you put the body back together when no one was sawn in two. It's more complex when there really is a body split in two, and mirrors and smoke don't do the trick.

Snake Oil and Nerd Harder

The enchantment of AI extends beyond anthropomorphic projections and magical marketing into an even more seductive trope: the belief that AI can solve virtually any problem we face as a society.⁸¹ This is techno-solutionism supercharged by claims that complex social, political, and economic challenges are merely engineering problems in disguise, waiting for the right technological intervention to make them vanish.⁸² No negotiation, no compromise, just the promise that code can optimize away inefficiencies, biases, and inequalities that have persisted throughout human history.

This myth operates through two distinct but complementary dynamics. The first is the familiar snake-oil marketing practice, whereby technology vendors actively promote their innovations as transformative solutions, promising efficiency, optimization, and progress through code. Companies market AI as the answer to everything from healthcare inefficiencies to educa-

tional inequality, often with claims that far exceed the capabilities of their systems.⁸³

But equally important is what some call the nerd harder phenomenon⁸⁴—when policymakers, faced with intractable social problems, turn to technology not because vendors are pushing solutions, but because they desperately want to believe that engineering can succeed where politics has failed.

The nerd harder mindset reflects a deeper willingness to blindly trust technological solutions that transcends the marketing of any particular vendor.⁸⁵ It's the belief that if we just apply enough computational power, data analysis, and algorithmic sophistication to social problems, we can engineer our way past the seemingly intractable challenges of human society. Such thinking is seductive for policymakers who are frustrated by the slow pace of democratic deliberation and the difficulty of building consensus around complex issues,⁸⁶ and who find the allure of clean, clickable solutions hard to resist.

Complex issues such as poverty, educational gaps, or political polarization are reimagined as technical malfunctions with information gaps to be filled, behaviors to be nudged, and inefficiencies to be smoothed. Healthcare becomes a data processing challenge, criminal justice becomes a prediction problem, and disinformation becomes a content moderation task. The complexity that makes scenarios genuinely difficult to improve—conflicting values, competing interests, historical injustices, and structural inequalities—gets abstracted away in favor of algorithmic formulations.

Consider how this logic recasts urban inequality. Rather than address housing policy or labor market precarity, smart cities promise to ease congestion, forecast crime, and streamline services.⁸⁷ The deeper roots of injustice remain, while the dashboard glows green.⁸⁸

This is not wishful thinking or naivety: it's politically convenient. Technology-driven solutions offer political advantages that traditional policy interventions do not: they can be implemented quickly without extensive stakeholder consultation, they avoid contentious debates about resource redistribution, and they create the impression of innovative leadership without requiring awkward political compromises.

The Automation Illusion

Central to techno-solutionist thinking is the “automation illusion”—the belief that complex human judgments can be reduced to algorithmic procedures without losing essential qualities that make those judgments valuable.⁸⁹ The illusion takes on new power when draped in the mystique of AI, whose pattern-finding powers seem to border on the supernatural.

It manifests most clearly in attempts to automate decisions that involve significant moral, cultural, or contextual dimensions. Essentially, code reveals its limitations where nuance is required.

Take, for example, efforts to automate hiring decisions through AI screening tools. These systems are marketed as capable of evaluating candidates through objective algorithmic analysis, eliminating human bias and inefficiency. The sales pitch is obvious: replace sub-

jective, potentially biased human judgment with neutral, data-driven decisions. However, in doing so, such tools reduce multidimensional human evaluation to tidy rankings.⁹⁰ The result? Metrics trump intuition, and potential gets reduced to pattern-matched proxies. Worse, bias is frequently still part of the equation.⁹¹

Welfare systems illustrate a similar tension. Here, automation is framed as efficiency: faster fraud detection, more accurate eligibility checks, and fewer human errors. But the reality is often different. Australia's Robodebt scheme⁹² and the Netherlands' "Toeslagenaffaire" (childcare benefits scandal)⁹³ are prime examples of how things can go horribly wrong. The Robodebt scheme refers to an Australian automated system using flawed income averaging calculations that wrongly accused hundreds of thousands of welfare recipients of owing debts, causing significant financial hardship, distress, and contributing to suicides.⁹⁴ In the "Toeslagenaffaire," the Dutch tax authorities used a self-learning algorithm to create risk profiles in an effort to spot child care benefits fraud. Thousands of families faced financial ruin, thousands of children were sent into foster care, and some victims committed suicide.⁹⁵ These disasters resulted from delegating complex assessments to an automated system without adequate testing, oversight, or consideration of its human impact, driven by a techno-solutionist belief in the efficiency and infallibility of automation. Context gets lost in translation. Rigid criteria misfire. The system may be logical, but it isn't always just.

Such considerations did not scare off the United Arab Emirates (UAE). Aiming to become a global leader in AI, the UAE is actively integrating AI into its judicial system. Initiatives include utilizing AI for case analysis—especially where outcomes are deemed clear-cut and do not require human discretion—such as document translation, summarization, analysis, case management, interactive case filing, and even predicting judicial outcomes based on historical data.⁹⁶

Educational technology provides another interesting example. Adaptive learning systems promise to personalize education by algorithmically adjusting content and pacing to meet the individual needs of each student. While these systems can certainly provide value, they often reduce learning to measurable skill acquisition while missing the relational, creative, and contextual dimensions of education that human teachers provide. What emerges is efficient instruction, not transformative teaching.⁹⁷

The automation illusion becomes even more problematic when it encounters what social scientists refer to as “tacit knowledge”: the type of understanding that humans possess but cannot easily articulate or formalize.⁹⁸ Medical diagnosis, for instance, involves more than pattern recognition in symptoms and test results, it also requires subtle intuitions about patient behavior, cultural context, and clinical experience that resist algorithmic capture.⁹⁹ When AI diagnostic tools are presented as superior to human judgment, they often excel in specific, well-defined tasks while missing the

broader contextual understanding that experienced practitioners bring to complex cases.

This illusion gains power from the genuine capabilities of AI systems in narrow domains. Machine learning excels at finding statistical patterns in large datasets, and such capabilities can genuinely enhance human decision-making in many contexts.¹⁰⁰ The problem arises when success in specific applications gets generalized into broader claims about AI's problem-solving potential, leading to the assumption that any domain involving data analysis or pattern recognition is suitable for algorithmic automation. Spoiler alert: that's not how things work.¹⁰¹

Democracy Through Algorithms

Perhaps nowhere is techno-solutionist thinking more seductive—or more dangerous—than in applications to democratic governance itself. The promise of “digital governance” suggests that technology can solve fundamental challenges of democratic participation, representation, and decision-making that have persisted throughout political history.

Digital platforms, such as vTaiwan¹⁰² in Taiwan and Decidim¹⁰³ in Barcelona, Spain, promise to enhance democratic participation through online deliberation tools that aggregate citizen input, facilitate discussion, and identify areas of consensus. They use sophisticated algorithms to organize discussions, highlight key arguments, and synthesize diverse viewpoints into actionable policy recommendations. The appeal is obvious as they suggest users can overcome the limitations of traditional democratic institutions through technolog-

ical innovation. They are the latest shiny tools to solve ancient problems.

Estonia's digital government initiatives showcase what this vision can look like at scale.¹⁰⁴ Citizens vote online, access government services, and participate in policy discussions through digital interfaces. The Estonian model is hailed by many as an example of frictionless governance, and a clean contrast to sluggish bureaucracies elsewhere.

Yet these initiatives, however innovative, struggle with fundamental tensions in democratic governance that technology cannot resolve. Online participation systems often exacerbate existing inequalities, amplifying the voices of those with access to technology and the necessary skills, while marginalizing others. Digital platforms can facilitate discussion, but they cannot resolve underlying disagreements about values, priorities, and resource allocation that make politics inherently contentious. They might widen access, but they do not level power.

More troubling, algorithmic mediation of political processes can subtly shift power from elected representatives to the engineers and designers who create the systems. When algorithms determine whose voices get heard, how discussions are organized, and what consensus emerges from deliberation, technical design choices become political decisions—and they are often made by unelected technical teams.

The blockchain voting movement exemplifies techno-solutionist thinking applied to electoral democra-

cy. Proponents argue that blockchain technology can solve problems of election security, voter access, and result verification through cryptographic protocols that eliminate the need for trust in human institutions.¹⁰⁵ The promise is compelling: replace fallible human election administration with mathematically guaranteed integrity. But blockchain voting systems, despite their technical sophistication, cannot address the social and political dimensions of electoral legitimacy. Elections derive their authority not just from technical security but from public trust in the institutions that conduct them.¹⁰⁶ When voters cannot understand or verify the technical systems that count their votes, even perfectly secure cryptographic protocols may fail to provide the public confidence that democracy requires.

Content moderation on social media follows the same techno-solutionist script.¹⁰⁷ These systems use machine learning to identify hate speech, misinformation, and policy violations across billions of posts daily. They promise to create safer, more civil online spaces through neutral application of community standards. Yet content moderation algorithms consistently struggle with context, cultural nuance, and the inherent ambiguity of human communication. Attempts to encode complex social norms into algorithmic rules often produce decisions that are technically correct but socially inappropriate, leading to controversies about censorship, bias, and cultural imperialism when global platforms apply uniform standards across diverse cultural contexts.¹⁰⁸

Moreover, when private companies make policy decisions that affect public discourse, the result is a form of

“platform democracy” whereby corporate employees make quasi-governmental decisions without the associated democratic accountability. The promise of neutral algorithmic governance thus paradoxically concentrates political power in the hands of unelected technical elites.

Smart Cities and Optimized Society

The smart city movement represents perhaps the most comprehensive expression of techno-solutionist thinking, by aiming to solve urban challenges through the comprehensive deployment of sensors, algorithms, and data analytics. Smart city initiatives promise to optimize traffic flows, reduce energy consumption, predict infrastructure needs, and enhance public safety through real-time data collection and algorithmic analysis.

Sidewalk Labs’ now-canceled Toronto waterfront project was a case in point.¹⁰⁹ The Google subsidiary¹¹⁰ envisioned a neighborhood optimized by data with everything tracked and tuned. From trash collection to foot traffic, sensors would feed a continuous feedback loop, creating an urban environment that could adapt in real-time to resident needs and preferences. It was seductive, and it was simplistic. However, real cities are messy. They contain competing priorities, unequal access, and historical scars that no algorithm can erase. Optimizing traffic flows might benefit drivers while harming pedestrians and cyclists. Predictive policing algorithms might reduce crime statistics while exacerbating racial tensions. Energy efficiency measures may lower costs, but they can also create uncomfortable living conditions for vulnerable populations.

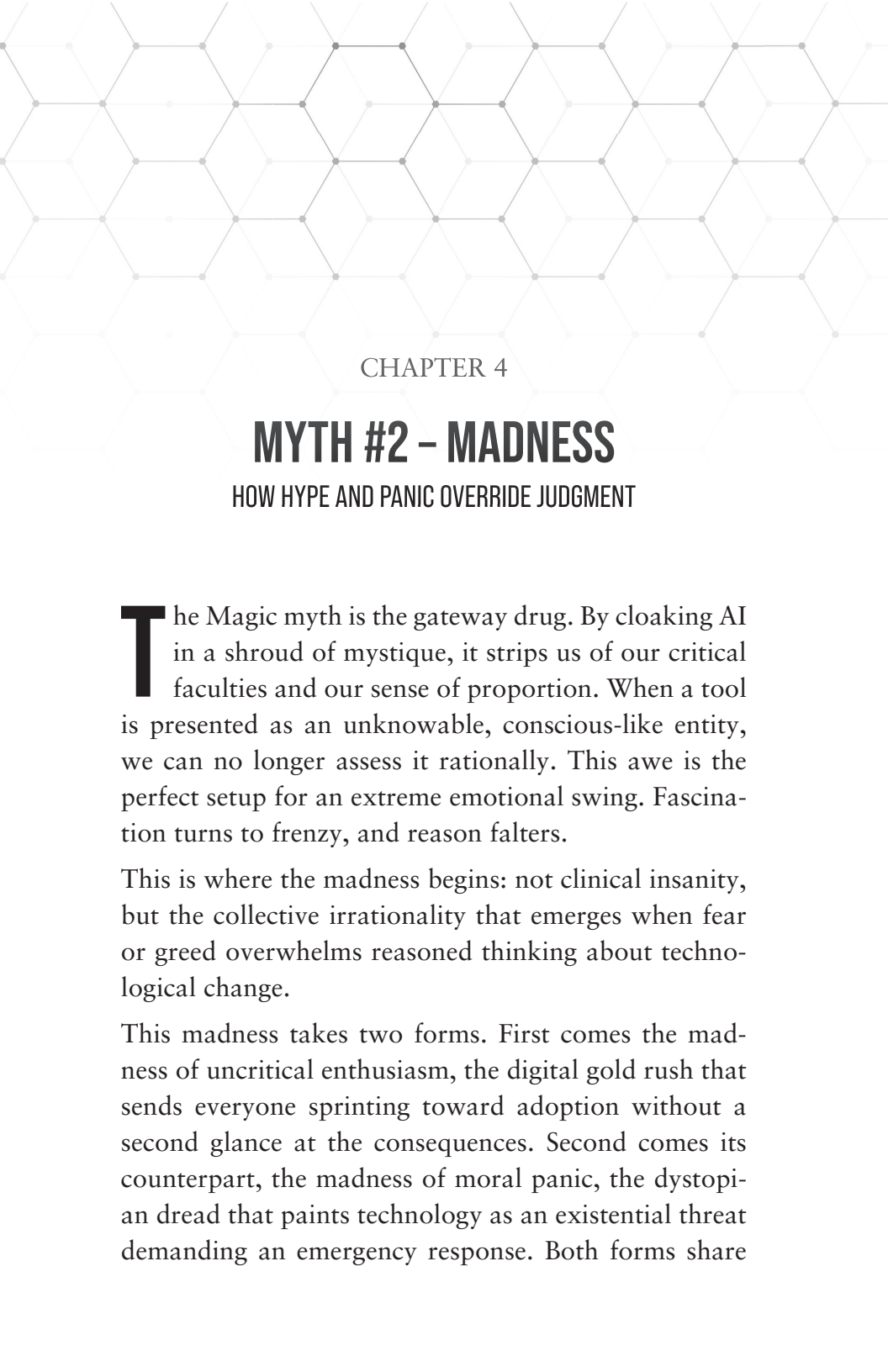
The Sidewalk Labs project ultimately collapsed partly because residents and activists questioned who would control the data the system collected and how algorithmic decisions would be made accountable to democratic oversight. The promise of optimization clashed with the reality that urban governance involves making choices between competing values that cannot be resolved through technical means alone.

Barcelona's smart city initiatives provide a more nuanced example of both the potential and limitations of technological approaches to urban governance.¹¹¹ It utilizes smart tools selectively to monitor air quality, manage waste, and reduce energy use. The systems have generated genuine improvements in municipal efficiency and environmental performance. But Barcelona's experience also illustrates the importance of avoiding panopticon solutions by maintaining democratic oversight over technological systems. The city has explicitly rejected comprehensive surveillance approaches, instead focusing on specific applications that clearly serve public goals while maintaining citizen privacy and democratic accountability. This more limited approach achieves real benefits while avoiding the totalizing logic of comprehensive optimization.

The narrative of the Chinese social credit system represents the logical extreme of smart city thinking applied to social governance. In Western discourse, it functions as a powerful cautionary tale in which optimization becomes surveillance, and public life is governed not by law but by a single, comprehensive score.¹¹² The reality, as researchers note, is more complex and fragment-

ed—less a single, unified system than a patchwork of dozens of local pilot programs and separate blacklisting systems, many of which focus on corporate compliance rather than individual social behavior.¹¹³ Yet, even in this fragmented, real-world application, the system demonstrates the core logic of techno-solutionism. Its stated promise is to create a more “trustworthy” society by monitoring citizen behavior and providing algorithmic scores that affect access to services and social benefits. Proponents argue that it optimizes social cooperation by creating incentives for good behavior.

This logic, whether applied monolithically or in fragments, demonstrates how algorithmic optimization can conflict with fundamental values, including privacy, autonomy, and human dignity. When algorithms are tasked with monitoring and judging human behavior comprehensively, what begins as efficiency can end in control.



CHAPTER 4

MYTH #2 – MADNESS

HOW HYPE AND PANIC OVERRIDE JUDGMENT

The Magic myth is the gateway drug. By cloaking AI in a shroud of mystique, it strips us of our critical faculties and our sense of proportion. When a tool is presented as an unknowable, conscious-like entity, we can no longer assess it rationally. This awe is the perfect setup for an extreme emotional swing. Fascination turns to frenzy, and reason falters.

This is where the madness begins: not clinical insanity, but the collective irrationality that emerges when fear or greed overwhelms reasoned thinking about technological change.

This madness takes two forms. First comes the madness of uncritical enthusiasm, the digital gold rush that sends everyone sprinting toward adoption without a second glance at the consequences. Second comes its counterpart, the madness of moral panic, the dystopian dread that paints technology as an existential threat demanding an emergency response. Both forms share

a dangerous trait in that they short-circuit democratic deliberation by appealing to emotional extremes and knee-jerk reactions rather than evidence-based analysis.

The Digital Gold Rush

Doomers vs. Boomers

The madness of the AI discourse is sustained by a theatrical and deeply misleading conflict between two seemingly opposed tribes: the doomers and the boomers.

The doomers are concerned with the speed of AI development. They speak in the language of existential risk, popularizing statistical-sounding buzzwords like “P(doom)” (probability of doom) to frame their fears of a machine uprising that could “kill us all.”¹ This apocalyptic framing, amplified by organizations such as the Future of Life Institute² and the Center for AI Safety,³ treats AI as a potential extinction-level event, on par with pandemics and nuclear war, prompting the question of whether we are “Ready for it?” This apocalyptic framing was perfectly captured by the September 2025 launch of the so-called “Doom Bible,” *“If Anyone Builds It, Everyone Dies,”* by the Machine Intelligence Research Institute (MIRI)⁴ researchers Eliezer Yudkowsky and Nate Soares. The book’s launch was a massive media spectacle, featuring billboards that read “We Wish We Were Exaggerating” and primetime TV appearances, all advocating for a global halt to AI development to prevent human extinction. The book’s proposed solutions were so extreme, including the bombing of data centers, that numerous media review-

ers dismissed the work not as science, but as “impractical”, “science fiction”, and “third-rate sci-fi.”⁵ The authors’ reactions in interviews also demonstrate how doomerism sidelines present-day concerns: when one author was asked about the immediate risk of centralizing power, he dismissed the concern, stating, “That really seems like the sort of objection you bring up if you don’t think everyone is about to die.”⁶

The boomers, on the other hand, preach a gospel of relentless progress. They are the techno-optimists who see AI development as not just beneficial but inevitable, and a force that will unlock untold prosperity and solve humanity’s greatest challenges.

On the surface, these two factions appear to be in opposition. But this apparent tension is mostly a distraction.⁷ Doomerism and boomerism are not truly opposing forces; they are two sides of the same coin, both minted in the same Silicon Valley furnace. Both deeply believe that Artificial General Intelligence (AGI) is inevitable and are equally fixated on it. AGI refers to an AI system with the ability to understand, learn, and apply intelligence across a wide range of tasks—essentially, machines with cognitive capabilities on par with, or exceeding, human beings.⁸ Unlike the specialized or “narrow” AI we use today, which excels at singular functions such as recommending movies or transcribing speech, AGI is envisioned as adaptable and self-improving across domains.⁹

This shared faith or apprehension, depending on the perspective, is animated by two underlying ideologies that are equally extreme. The doomer camp gave rise

to Effective Altruism (EA), a movement that originated in Oxford philosophy circles and was subsequently bolstered by funding from tech billionaires.¹⁰ EA advocates for using extreme rationality to do the most good, and under the influence of thinkers such as Nick Bostrom, a philosopher, author notably of *Superintelligence* and *Deep Utopia*, and founder of the Future of Humanity Institute, it identified rogue AI as a primary existential threat.

In reaction, some of the more vocal voices in the boomer camp rallied around a counter-ideology called Effective Accelerationism (e/acc).¹¹ What started as a joke to lampoon EA quickly enshrined its polar opposite spirit, elevating the maximalist view of flooring the accelerator on both the development and adoption of AI.

Yet, despite their apparent opposition, both camps serve the same end. They ensure the debate remains centered on a speculative, sci-fi future, effectively sidelining present-day concerns about labor exploitation, algorithmic bias, and environmental impact, both in terms of how they shape discourse and how they fund research. This is where the concept of “criti-hype” becomes so useful. Coined by Lee Vinsel, it is a form of criticism that takes its subjects’ claims to genius at face value and, paradoxically, inflates the power of the technology it critiques.¹² When organizations release open letters calling for a pause on “giant AI experiments” out of fear of existential risk, they are not truly challenging the hype, rather they are amplifying it. They reinforce the boomers’ premise that the technology is godlike in its potential. This plays into companies’ hands by portray-

ing them as creators of otherworldly systems, a more favorable image than being seen as vendors of products that, at this stage, are sometimes flawed and still under development.¹³

The framing allows a small group of unelected tech leaders to dictate the terms of their own regulation, ensuring that when policymakers do act, they are focused on the “frontier models” that only a handful of companies can build. The fight between doomers and boomers is not a genuine debate about the future of technology; it is more akin to political theater that ensures the existing empires of a handful of tech companies remain in control.

AI and the Art of Selling Shovels

When ChatGPT showcased its seemingly magical capabilities in late 2022, it triggered a powerful psychological phenomenon. The event created what psychologists call the “availability heuristic,” a cognitive bias that makes us give undue weight to recent, vivid examples when making judgments.¹⁴ Suddenly, every small advance in machine learning felt like proof that the technological singularity was just around the corner. Such a reaction illustrates Amara’s Law: “We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.”¹⁵

The collective overestimation sparked a stock market frenzy, driving financial markets to soar since 2023. When the media company BuzzFeed announced it would use OpenAI’s technology to generate content, its long-sagging stock skyrocketed 150% in a day.¹⁶ Investors did not wait to see results—the mere sprinkle

of AI pixie dust caused a mini gold rush on the NASDAQ. The pattern repeated itself with almost comical predictability as companies began adding “AI” to their names and watched their valuations soar, echoing the dot-com bubble’s craze for anything with an “e” prefix. The mere suggestion of a link to AI was enough to transform corporate fortunes overnight.

The gold rush metaphor is particularly apt. In the California Gold Rush of 1849, very few prospectors struck it rich. The people who consistently profited were those selling supplies to miners, such as Levi Strauss with his durable denim, merchants selling pickaxes, and companies running the supply routes. Their success embodied the timeless wisdom famously established by Sam Brannan, California’s first millionaire: “During a gold rush, sell shovels.”¹⁷ The AI boom followed identical dynamics. While AI startups burned through venture capital trying to build the next ChatGPT, Nvidia became the most valuable shovel manufacturer in history.

Jensen Huang’s company didn’t need to promise AGI or claim it would solve world hunger. Nvidia simply made the Graphics Processing Units (GPUs), the computational bedrock for training AI models. While everyone else hunted for digital gold, Nvidia’s stock grew by more than 2,000% between 2022 and 2024. The company’s market capitalization, which exceeded \$3 trillion in 2024—making it more valuable than the GDP of many countries¹⁸—continued its historic climb. In October 2025, it became the first company in history to briefly reach a \$5 trillion valuation.¹⁹

The market-level gold rush was a mirror of the deep competitive frenzy raging within the AI labs themselves. Nowhere was this more apparent than at OpenAI, whose internal philosophy was built on a near-apocalyptic assumption of the need to be first or perish. As Karen Hao documents in *Empire of AI*,²⁰ this belief set a “ticking clock on each of [the] research advancements,” justifying the consumption of unfathomable resources and the premature release of powerful technologies. Such anxiety was not abstract, it was personal and specific. It was initially fueled by an intense rivalry with Google’s DeepMind and later magnified by the breakaway of its own top researchers to found the rival lab Anthropic. A constant fear of being outshined created a perpetual state of corporate anxiety, ensuring that madness was not just a feature of the stock market, but a part of the core operating system of one of the vocal architects of the hype.

The Market Bubble and the Backlash

The digital gold rush culminated in the AI market correction of August 2025, which perfectly captured the violent swing from manic hype to sudden panic. A frenzy had been building for more than a year, but the pin that pricked the bubble was not a dramatic technological failure, but a dose of sober reality from a seemingly unlikely source: an MIT-labeled report.²¹

The report asserted that a staggering 95% of enterprise generative AI pilot projects were failing to deliver on their promises of rapid revenue growth. While companies had poured billions into AI initiatives, the much-

hyped revolution was, for the most part, not materializing on the bottom line.

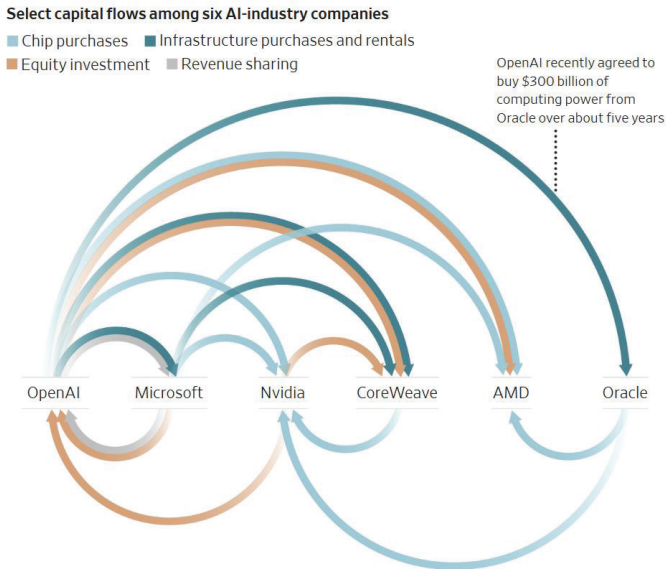
Adding fuel to the fire, OpenAI's CEO, Sam Altman, publicly acknowledged the possibility of an AI bubble just as his company's much-anticipated GPT-5 release underwhelmed users. The confluence was toxic. The MIT report provided the data, and Altman's comment provided the rationale. Spooked investors, suddenly terrified of being left holding the bag, triggered a massive tech selloff. The S&P 500 shed \$1 trillion in value, a staggering loss driven not by a drastic change in the technology's potential, but by the sudden collapse of the collective narrative.²² In a world of market madness, nuance is the first casualty. The AI rush had given way to panic, demonstrating that when it comes to technology, the distance between euphoria and fear is dangerously short.

Of course, the story wasn't that simple. Rebuttals to the MIT report quickly emerged, arguing that any perceived "failures" were merely the natural growing pains of a transformative technology.²³ Some of its findings, however, were confirmed by McKinsey's "State of AI in 2025" report.²⁴ It found that while a Digital Gold Rush had led nearly 90% of companies to adopt AI, "nearly two-thirds" had failed to scale it across the enterprise, and only 39% could report any enterprise-level bottom-line impact.

Many observers reflexively reach for the Gartner Hype Cycle to explain this phenomenon, with its familiar arc from a "Peak of Inflated Expectations" to a "Trough of Disillusionment."²⁵ However, the model is a poor

fit for AI. The field does not follow a neat progression toward productivity. Instead, its history is one of alternating “springs” of intense funding and hype, followed by “winters” of stagnation and drying funds.²⁶ Understanding this cyclical pattern is crucial because it reveals that the madness is not a one-time event but a recurring feature of the AI landscape.²⁷

The scale of this particular gamble, however, has reached unprecedented proportions. By 2025, the U.S. economy had become, as one analyst described it, “one big bet on AI.”²⁸ The figures from the first half of the year were staggering: investment in AI-related information processing was reportedly responsible for 92% of all U.S. GDP growth, with the rest of the economy growing at a mere 0.1%. In the stock market, AI companies accounted for 80% of all gains in 2025. This led The Guardian to declare the “AI valuation bubble is now getting silly,” comparing it to the “irrational exuberance” of the 1996 dot-com bubble.²⁹ The Guardian highlights an unprecedented “concentration risk,” with the “Magnificent 7” tech companies—representing the leading AI companies—accounting for over a third of the S&P 500. This is compounded by correlation risk from circular financial deals—such as Nvidia investing in OpenAI, which in turn buys Nvidia’s chips—a structure that critics argue ensures capital is misallocated.³⁰

Figure 3—Select capital flows among six AI companies³¹

Note: Some chip purchases are through intermediaries. Some investments and other arrangements subject to conditions.

Sources: staff reports; Morgan Stanley

This raises the stakes of the Madness myth far beyond a simple tech selloff, as the entire economic future of the U.S., and by extension the rest of the world, might depend on the fate of the hype—a precarity put on full display in early November 2025, when global stock markets fell sharply, driven by widespread fears that the AI bubble was finally bursting.³²

The FOMO Policy Syndrome

Madness is not confined to financial markets; it echoes loudly in the corridors of power. Governments have rolled out the red carpet for AI firms, sometimes relaxing regulations in their hurry to attract investment.

The United Kingdom's AI White Paper, published in March 2023, pointedly avoided creating new rules, instead suggesting that existing regulators could figure out AI governance on their own.³³ The rationale? Don't stifle innovation with premature rules.

France positioned itself as the "startup nation" for AI, offering tax breaks and streamlined visa processes for AI researchers.³⁴ Singapore launched a billion-dollar AI initiative, despite having a population smaller than that of most major cities.³⁵ The city-state's leadership explicitly framed the investment as necessary to avoid being left behind in the AI race.

Policy based on fear of missing out (FOMO) creates competitive deregulation in a race to offer the most business-friendly environments. The result is a race to the bottom where meaningful oversight becomes politically difficult, if not impossible.

Such a mindset is evident in the rhetoric of an arms race, which transforms technological development into a geopolitical contest. When Russian President Vladimir Putin declared in 2017 that "whoever leads in AI will rule the world,"³⁶ he identified a competitive dynamic that has since shaped national AI strategies. China's response to Google-built AlphaGo's victory over Go champion Lee Sedol is a perfect illustration. The match was immediately labeled China's "Sputnik moment,"³⁷ sparking a massive national AI program with clear goals: lead the world in AI theory by 2025 and achieve major breakthroughs by 2030.

U.S. policymakers responded with predictable urgency. The 2018 National Defense Authorization Act established the Joint Artificial Intelligence Center (now integrated into the Chief Digital and Artificial Intelligence Office) specifically to ensure military AI superiority.³⁸ The 2021 National Security Commission on Artificial Intelligence warned that the United States faced an “AI winter” if it did not dramatically increase investment and remove regulatory hurdles.³⁹

The race reached new heights in late July 2025, when the United States and China released parallel AI action plans within three days of each other. The timing may have been a coincidence, but the content exemplified the competitive madness. The Trump administration’s plan, titled “Winning the Race,”⁴⁰ is a masterclass in policy FOMO. It explicitly frames AI as a “zero-sum global competition” and declares that the US and its allies must “win this race” to achieve “global dominance.” True to the principles of competitive deregulation, the plan abandons the previous administration’s safety-first approach in favor of an innovation-first strategy, vowing to repeal “burdensome regulations” and to pressure allies to align with its *laissez-faire* model.⁴¹

China’s Global AI Governance Action Plan, announced at the World AI Conference in Shanghai, played a different tune but was part of the same orchestra.⁴² Where the US plan is overtly adversarial, Beijing’s is rhetorically collaborative, emphasizing an inclusive, multilateral approach via the United Nations and positioning China as a voice for the Global South. Yet, this coop-

erative language masks a familiar ambition. While the United States seeks dominance through market power and a proprietary tech stack, China aims to embed its own state-centric governance model into global standards through strategic diplomacy.⁴³ The result is that one nation is racing to tear down the guardrails, while the other is racing to build the road. Both are sprinting, driven by the same fear of being left behind, and in the process, ensuring the madness goes global.

The problem with governments treating AI development like an arms race is that they misunderstand the nature of AI. Unlike nuclear weapons, AI capabilities are not a zero-sum game. When China develops better natural language processing, it doesn't eliminate the United States' ability to build similar systems. AI research is cumulative and often benefits more from open collaboration than from secretive competition. This reality is evidenced by the open-source nature of some parts of the industry, where foundational research papers are published immediately and many leading models are publicly available, allowing competitors to replicate and build upon each other's work rapidly. The arms race framing creates an artificial sense of scarcity, prompting policy responses rooted in competitive fear rather than collaborative opportunity. The analogy is especially misleading because the goal of the AI race is not to build a secret weapon to be held in reserve, but to deploy commercial technologies whose value increases with widespread use.⁴⁴

The Infrastructure Madness

While many focused on the software side of the AI revolution, a parallel surge was happening in the hardware world, driven by data center demand. Companies rushed to build massive facilities capable of housing thousands of GPUs, leading to a construction boom.

Microsoft alone planned to spend \$80 billion on data centers between 2024 and 2025,⁴⁵ with Amazon, Google, and Meta announcing similarly ambitious expansion plans. Once again, the shovel-selling analogy holds. While AI companies competed to build the most impressive models, real estate developers and construction firms profited from the infrastructure boom on the sidelines.

The surge in data centers strained local power grids⁴⁶ and sparked conflicts with communities over energy and water use, as well as broader environmental impacts. The energy implications, in particular, once overlooked, are now staggering and are completely rewriting national energy plans. The International Energy Agency (IEA) projects that by 2030, global data center electricity demand will more than double to 945 TWh, surpassing the entire current electricity consumption of Japan.⁴⁷ In the U.S. alone, 2024 data center consumption was already equivalent to the annual demand of Pakistan.⁴⁸ This insatiable demand is forcing a hard reversal on decarbonization goals, as utilities delay the retirement of coal-fired plants⁴⁹ and plan new natural gas facilities to meet the shortfall. In a stark illustration of this new reality, Microsoft is even reopening part of

the Three Mile Island nuclear power plant to provide carbon-free energy for its AI operations.⁵⁰

Ireland provides a cautionary tale. The country's favorable tax policies and cool climate made it attractive for data center investment, but by 2023, data centers accounted for more than 20% of Ireland's electricity consumption.⁵¹ The government was forced to impose a moratorium on data center construction in the Dublin area, effectively acknowledging that the infrastructure rush had outpaced the country's ability to provide sustainable power.

Yet, even as infrastructure constraints became apparent, policymakers worldwide continued to offer incentives for AI companies to locate in their jurisdictions. Intel received \$8.5 billion in subsidies to build semiconductor fabrication facilities in different U.S. states,⁵² while Taiwan Semiconductor Manufacturing Co. was granted \$6.6 billion for its U.S. operations in Arizona.⁵³ These investments were justified partly as necessary for AI competitiveness, but they also reflected a FOMO-driven bidding war between states and countries.

Moral Panic and Dystopian Fears

If the digital gold rush is the madness of uncritical enthusiasm, moral panic based on irrational fear is its twin. Whereas hype cycles push societies toward reckless adoption, moral panics drive them toward reactionary prohibition. Both forms of madness share the same problem in that they replace careful analysis with

emotional response, and the results are equally damaging.

The pattern is depressingly familiar. Throughout history, innovations have triggered moral panics that far exceeded any actual risks. Comic books were blamed for juvenile delinquency in the 1950s. Rock music was supposedly corrupting youth in the 1960s. Video games were demonized as training simulators for violence in the 1990s. Each panic followed the same script of dramatic warnings about civilization-threatening doom, calls for immediate government action, and eventually, the quiet recognition that the fears were wildly overblown.

Panic about AI follows this established playbook, but with a contemporary twist that makes the fear feel more potent. The sophistication of modern AI systems provides just enough genuine capability to make apocalyptic scenarios seem plausible, even when they remain highly speculative. This creates a sci-fi feedback loop, where fictional portrayals of AI in movies and books shape public perception, which in turn shapes policy responses to real-world technology.

The Deepfake Apocalypse That Wasn't

The election deepfake panic of 2019 and 2020 serves as an illustration of how technical possibilities can spiral into policy panic, resulting in legislation that addresses phantom problems while real issues continue to fester.

The script began predictably enough. As computer scientists demonstrated increasingly sophisticated face-swapping technology, academic papers warned of

an impending “epistemic apocalypse” where truth itself would become unknowable.⁵⁴

Central to these warnings was the clever and troubling concept of the “liar’s dividend.” Coined by U.S. law professors Robert Chesney and Danielle Citron, the theory posits that the most insidious threat is not that people will believe the fake content. Rather, it was that bad actors could use the mere possibility of a deepfake to discredit authentic evidence, dismissing genuine audio or video of their own misconduct as a sophisticated fabrication.⁵⁵ In this scenario, accountability evaporates in a fog of plausible deniability, eroding the shared reality on which democratic debate depends.

The media, naturally, amplified such fears. Headlines proclaimed that deepfakes would “wreak havoc on society”⁵⁶ and usher in an “age of paranoia.”⁵⁷ Each new technical demo seemed to confirm that democracy itself was under siege.

However, the panic narrative then collided with a more mundane reality. When researchers examined the online misinformation circulating during the 2020 election, deepfakes played a minimal role in election-related deception.⁵⁸ Old-fashioned manipulation, including the use of misleading statistics, out-of-context images, and straightforward false claims, remained far more effective. The disconnect was stark, and a comprehensive analysis revealed that more than 90% of deepfake videos online were non-consensual pornography targeting women, not political disinformation.⁵⁹ The real, immediate harm was not to election results, but to the lives of ordinary people, overwhelmingly women.

Despite this empirical vacuum, the policy response was both swift and sweeping. California rushed to ban deepfake political videos within sixty days of the elections,⁶⁰ while Texas criminalized deepfakes intended to harm candidates or influence elections.⁶¹ Congressional hearings treated the technology as equivalent to weapons of mass destruction,⁶² while intelligence agencies warned of foreign interference deepfake campaigns that never materialized at scale.

The Defense Advanced Research Projects Agency (DARPA) launched the \$68 million MediFor⁶³ and SemaFor⁶⁴ programs to develop deepfake-detection technology, treating the threat as a national security priority on par with biological weapons or cyber warfare. The goal was to combat “large-scale, automated disinformation attacks” that existed primarily in the realm of speculation.

The irony, of course, is that while policymakers were busy building defenses against a phantom menace, genuine threats began to emerge in ways the laws driven by panic had not anticipated. Slovakia’s parliamentary election of September 2023 marked a turning point when a deepfake audio recording surfaced just two days before the vote, allegedly capturing opposition leader Michal Šimečka discussing election fraud.⁶⁵ The fake audio emerged on Telegram and quickly spread to TikTok, YouTube, and Facebook, despite fact-checkers’ attempts to contain it.

In 2024, the global election cycle brought more examples of deepfake interference. In January, thousands of New Hampshire voters received AI-generated robocalls

in U.S. President Joe Biden's voice, urging Democrats not to vote in the state's primary.⁶⁶ Yet even as these first high-profile cases emerged, comprehensive analysis revealed that "the feared wave of deceptive, targeted deepfakes didn't really materialize."⁶⁷

The evolution from hypothetical threat to real-world tactic has since accelerated. In late October 2025, just days before Ireland's presidential election, a convincing deepfake audio clip of candidate Catherine Connolly circulated widely on social media, appearing to show her announcing her withdrawal from the race.⁶⁸

That same week, the elections in the Netherlands provided a stark example of a different, cruder AI tactic. Fake, AI-generated images of center-left leader Frans Timmermans were circulated on Facebook, depicting him being arrested and giving money to a Muslim couple.⁶⁹ The images, posted by members of a rival political party, functioned as modern propaganda, using AI not for high-tech deception but for low-brow character assassination.

In the United States, this trend shifted from fringe attacks to mainstream strategy. President Donald Trump's 2025 campaign fully embraced AI-generated content on his Truth Social platform. His account frequently posted AI-generated images depicting him in fantastical scenarios—as a crowned king, a Nobel Peace Prize winner, or a warrior.⁷⁰ This strategy used AI not to deceive voters but to entertain them and reinforce his persona as a larger-than-life figure. This tactic has been recently mirrored by opponents, with California Governor Gavin Newsom's press office using AI-generated imag-

es to mock President Trump in a similar, meme-driven style.⁷¹

This created a bizarre new political dynamic: politicians were simultaneously using AI as a propaganda tool while also signing legislation to regulate its “dangers.”⁷² The panic over deceptive deepfakes had missed the more immediate reality: AI was being laundered into the political mainstream as spectacle, blurring the lines between satire and propaganda.⁷³

The Children Are Always in Danger

Nothing ignites a tech moral panic quite like the cry to protect the children, and AI has been no exception. When ChatGPT launched in late 2022, schools across the United States reacted almost instantaneously, shutting down access before any systematic study had been proposed. The New York City Department of Education banned ChatGPT on every school device on January 10, 2023, warning that it could “undermine student learning,” despite lacking evidence of widespread misuse or clear harm.⁷⁴

Seattle Public Schools followed suit, blocking the tool on all district hardware to “require original thought and work from students.”⁷⁵ Within weeks, districts from Los Angeles to Baltimore issued similar bans, invoking fears of a cheating epidemic rather than exploring potential educational benefits.⁷⁶ This was a moral panic in miniature, and a commercial opportunity for others.

Detection tool vendors rushed to fill the panic-driven void. Turnitin activated its AI-writing detector in April

2023, claiming it could nearly flawlessly flag 97% of GPT-3 and ChatGPT submissions with a lab-tested false-positive rate of less than 1%.⁷⁷ By January 2024, it had scanned more than 200 million student papers, despite educators' warnings that its false positives could unjustly penalize students. The U.S.-born panic quickly crossed borders. In Australia, schools imposed similar bans mainly based on U.S. media coverage.⁷⁸

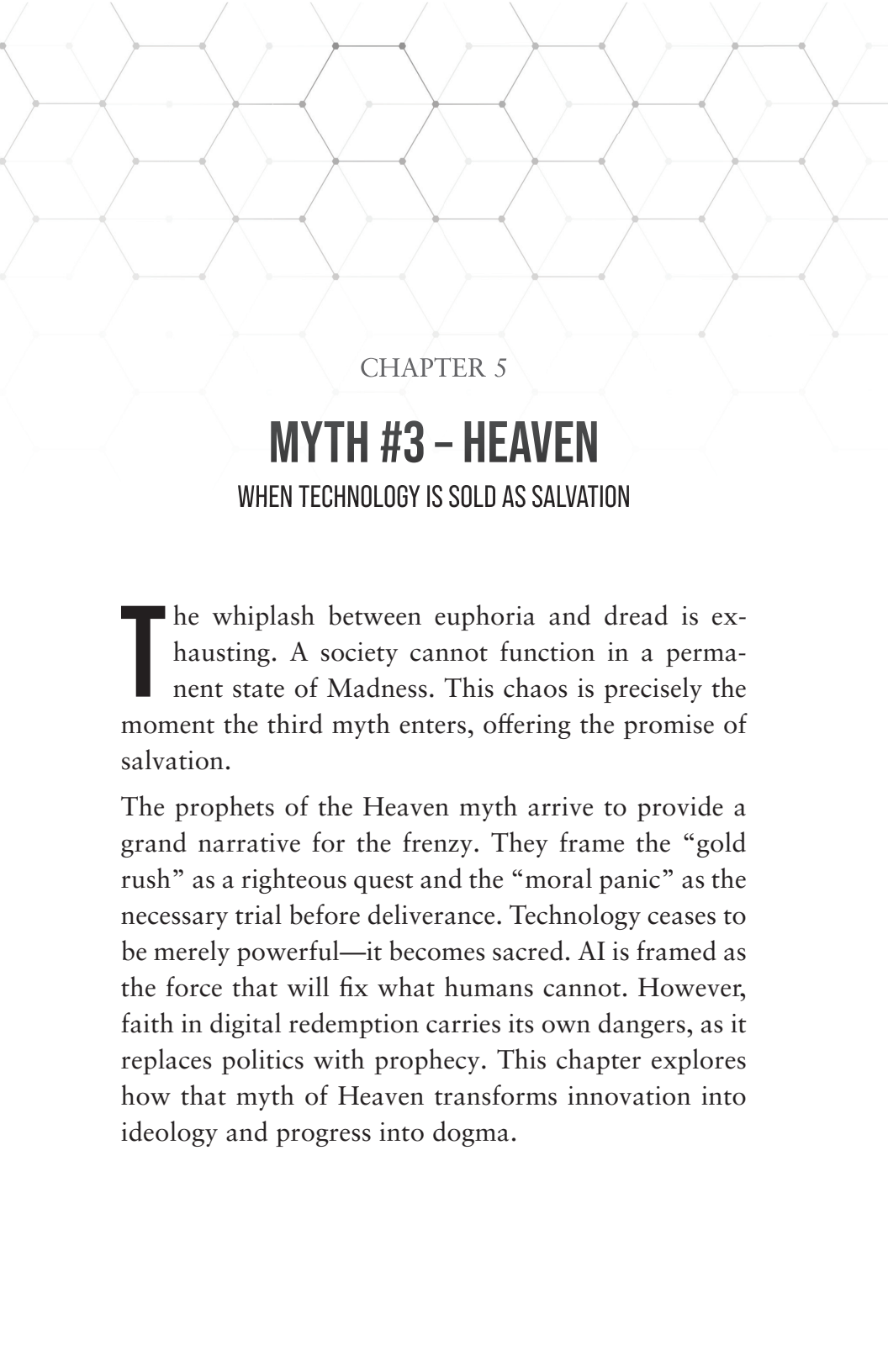
Then, as often, the researchers caught up, and they unpacked a much more nuanced picture. A randomized controlled trial at Stanford found that students using an AI-powered tutor learned more than twice as much in half the time as they did in traditional active-learning classes and reported a deeper conceptual understanding.⁷⁹ Surveys from other institutions showed that most students were using AI not as a ghostwriter, but as a partner for brainstorming, editing, and clarifying their own thinking.

It is an ancient fear and a familiar pattern. In *Phaedrus*, Plato—ironically, through the written word—warned that the invention of writing itself would “implant forgetfulness” in learners' souls. He feared they would come to rely on external marks rather than their own internal resources, achieving only the “semblance of wisdom” rather than true understanding.⁸⁰

In the 1970s and 1980s, pocket calculators were denounced as a crutch that would erode basic arithmetic skills; several U.S. states barred them from standardized tests.⁸¹ A similar panic erupted over video games after the 1993 U.S. Senate hearings on violent titles like *Mortal Kombat*, where senators warned of “graph-

ic gore” harming young minds despite scant evidence linking gameplay to real-world violence.⁸²

Across all these cases, the same script unfolds of hasty reactions, worst-case framing, solutions marketed by profit-seeking vendors, and legislation dashing ahead of balanced, empirical research.



CHAPTER 5

MYTH #3 – HEAVEN

WHEN TECHNOLOGY IS SOLD AS SALVATION

The whiplash between euphoria and dread is exhausting. A society cannot function in a permanent state of Madness. This chaos is precisely the moment the third myth enters, offering the promise of salvation.

The prophets of the Heaven myth arrive to provide a grand narrative for the frenzy. They frame the “gold rush” as a righteous quest and the “moral panic” as the necessary trial before deliverance. Technology ceases to be merely powerful—it becomes sacred. AI is framed as the force that will fix what humans cannot. However, faith in digital redemption carries its own dangers, as it replaces politics with prophecy. This chapter explores how that myth of Heaven transforms innovation into ideology and progress into dogma.

Algorithmic Redemption

If Chapter 3 revealed how we mythologize AI systems as magical and superhuman entities, this Chapter examines a deeper phenomenon: the transformation of AI into an object of quasi-religious faith. This technological theology doesn't merely mystify how AI works; it sanctifies what AI promises to become. Where the magical approach focuses on AI's immediate capabilities, salvation narratives project those capabilities into a future where technology transcends the messy, imperfect business of human governance entirely.

A theological dimension is clearly visible in the language of AI's most prominent evangelists; it is a language of prophecy, not product development. When Altman describes AGI as humanity's "final invention," he's not offering a technical roadmap but what some could interpret as a prophecy of technological rapture.¹ When Google's former CEO Eric Schmidt proclaims that AI represents "the most important thing that's going to happen in about 500 years, maybe 1,000 years in human society, and it's happening in our lifetime,"² skeptics could argue that sounds like salvation through silicon. When futurist Ray Kurzweil proclaims that AI will enable humans to surpass biological limitations, achieving something approaching immortality through technological merger, he sounds less like a researcher and more like a herald of a digital afterlife.³

Such framing creates what scholars term "techno-theodicy," which is the belief that any apparent problems with AI systems must serve higher purposes in an inevitable march toward digital paradise.⁴ Just as tradition-

al theodicy explains suffering as part of a divine plan, this dogma rationalizes algorithmic harms as necessary growing pains on the path to an automated utopia. Present inequalities, surveillance overreach, and the erosion of democratic norms become acceptable costs for a future transcendence that technology will, one day, deliver.

The salvation narrative operates on three core doctrines of faith. First, the belief in inevitable progress, which treats AI development as a force of nature rather than the result of a series of human choices, making resistance seem both futile and foolish. Moore's Law becomes moral law, and technological advancement represents not just possibility but obligation. Second, the belief in transcendent optimization whereby AI will eventually solve problems that have plagued humanity throughout history, from scarcity to mortality. Finally, the doctrine of necessary faith demands that society trust in AI's benevolent trajectory, even when the evidence of current harms suggests otherwise.

Together, these beliefs create a policy environment where democratic oversight seems not just unnecessary but actively counterproductive. If AI development is inevitable and its destination is paradise, then regulation is simply misguided interference. If AI will eventually transcend human limitations, then present problems pale in comparison. And if faith in AI's benevolence is required to realize its potential, then criticism becomes a form of technological heresy.

The Gospel of Innovation

The salvation narrative borrows heavily from prosperity theology,⁵ promising that faith in technological progress will be rewarded with material abundance and social advancement. Silicon Valley's venture capital ecosystem exemplifies this gospel, where disruptive innovation becomes a form of secular prayer and successful IPOs represent divine blessing on technological faith.

This prosperity theology manifests most clearly in the promise that AI will create unprecedented economic growth that benefits everyone. Founder of the World Economic Forum, Klaus Schwab, describes a Fourth Industrial Revolution in which AI-driven automation is not a threat to employment but a pathway to universal prosperity.⁶ The narrative promises that while AI may eliminate some jobs, it will create jobs that are more fulfilling, creative, and well-compensated than ever.

Yet this gospel consistently avoids difficult questions about distribution and power. If AI creates enormous wealth, who will control it? If automation increases productivity, who will benefit? If new technologies eliminate traditional employment, what social safety nets will support displaced workers during transition periods that may last decades? The salvation narrative treats these as technical problems that future innovation will solve rather than political questions requiring democratic deliberation.

The cryptocurrency movement illustrates how this prosperity theology operates in practice. Bitcoin and other digital currencies were initially promoted not

just as alternative monetary systems but as pathways to financial liberation that would free individuals from government control and traditional banks.⁷ Early evangelists promised that cryptocurrency adoption would democratize finance, eliminate economic inequality, and create prosperity.

The reality proved more complex. While some early adopters achieved significant wealth through cryptocurrency speculation, the broader promises of financial democratization remained largely unfulfilled. Instead of eliminating traditional power structures, cryptocurrencies often replicated them: wealthy individuals and institutions accumulated the largest holdings, energy-intensive mining operations concentrated in regions with cheap electricity, and market manipulation by large players created volatility that harmed smaller investors.

More troubling, the prosperity gospel of cryptocurrency led to policy environments where regulatory oversight was dismissed as interference with inevitable technological progress. When lawmakers proposed consumer protections or market regulations, cryptocurrency evangelists characterized them as attacks on innovation rather than legitimate democratic efforts to prevent fraud and protect the most vulnerable.

The Doctrine of Algorithmic Objectivity

Another core tenet of the Heaven myth is faith in Algorithmic Objectivity—the doctrine that algorithms are not just capable (as the Magic myth claims), but inherently better and purer than flawed human judgment.⁸ This belief holds that because AI optimizes for efficien-

cy and consistency, it naturally eliminates human failings such as bias and corruption, thereby promising inherently more just and effective outcomes by removing humans from the loop.

Perhaps the most literal illustration of this faith in algorithmic purity is Albania's appointment of a virtual AI "minister" in September 2025. Prime Minister Edi Rama tasked the AI, named "Diella," with the explicit goal of fighting corruption by overseeing public tenders and procurement. The Prime Minister's claim that Diella would ensure "public tenders will be 100% free of corruption" embodies the doctrine of algorithmic objectivity, trusting a complex, deeply human political problem to an algorithm in the hope of achieving a purity that human institutions have failed to deliver.⁹

This doctrine also transpires in discussions of AI applications in criminal justice, where proponents argue that algorithmic risk-assessment tools will reduce bias and enhance public safety by replacing subjective human judgment with objective statistical analysis.¹⁰ The theory suggests that judges and parole officers, despite their experience and training, introduce bias and inconsistency that mathematical models can eliminate. In this case, the analysis rests on a misunderstanding of both algorithms and justice. Algorithmic systems reflect the biases present in their training data and in how the models are built, often amplifying discriminatory patterns. When predictive policing risk assessment tools such as PredPol¹¹ and the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)¹² show higher false positive rates for Black defendants, this

represents not algorithmic objectivity but mathematical encoding of historical discrimination.¹³ Algorithmic objectivity mischaracterizes justice as an optimization problem rather than an ongoing process of democratic deliberation about competing values. Criminal justice is much more than predicting recidivism. It involves balancing public safety against individual liberty, rehabilitation against deterrence, and efficiency against due process. An algorithm cannot resolve value conflicts because they represent disagreements about social priorities that require democratic resolution.

The doctrine extends to education, healthcare, employment, and social services, consistently promising that algorithmic decision-making will improve outcomes by eliminating human limitations. Educational AI will personalize learning more effectively than human teachers. Healthcare AI will diagnose diseases more accurately than human physicians. Employment AI will identify qualified candidates more fairly than human recruiters.

Each application follows the same faith-driven pattern: identify limitations in current human-centered processes, promise that algorithmic alternatives will eliminate them, and frame resistance to automation as opposition to progress itself—pure heresy.

Digital Manifest Destiny

The salvation narrative sometimes also incorporates elements of Manifest Destiny thinking,¹⁴ treating technological expansion as a natural right that justifies displacing existing institutions and practices. Just as nineteenth-century belief in Manifest Destiny portrayed westward expansion as the inevitable expression of

U.S. values, Digital Manifest Destiny treats technological adoption as a natural evolution that serves a higher purpose, regardless of the immediate costs.

Digital expansionism appears clearly in discussions of AI governance, where Silicon Valley leaders consistently argue that U.S. technological leadership serves global interests because U.S. values of innovation and entrepreneurship will guide AI development in beneficial directions.¹⁵ Schmidt's National Security Commission on Artificial Intelligence exemplified this thinking by framing AI competition with China as a struggle between democratic and authoritarian values.¹⁶ While legitimate concerns exist about authoritarian applications of AI surveillance and control, Schmidt's framing suggested that U.S. technological dominance naturally serves democratic purposes, regardless of how AI technologies are actually deployed.

The narrative suggests that concentrating AI development in U.S. companies represents not corporate capture of public technology, but a natural expression of democratic values that will benefit humanity universally. It conveniently obscures how U.S. companies have sometimes pursued projects that enable authoritarian practices when doing so aligns with political or commercial interests. Google's Project Dragonfly, which aimed to create a censored search engine for the Chinese market, and Amazon's marketing of its facial-recognition system, Rekognition, to law enforcement, both reveal how moral narratives can bend to expediency. Similarly, Microsoft's partnerships with Immigration and Customs Enforcement (ICE) demonstrate

how AI tools can reinforce surveillance at home, not just abroad.

These examples underscore a long-standing tension between the rhetoric of technological salvation and the realities of power. The U.S. has often cast its technological expansion as a civilizing or liberating force, yet such narratives have historically served national and corporate interests rather than universal democratic ideals. The nineteenth-century zeal of Manifest Destiny found new expression in the twenty-first-century Silicon Valley's proclaimed belief in progress through innovation. But as with its predecessors, the question remains: who is being "saved," and at whose expense?

The Altar of AGI

These salvation doctrines all converge on a single, ultimate promise (though some would see it more as a threat): the arrival of AGI. This is the promised land, the point on the horizon where today's narrow, task-specific systems evolve into something with human-like cognitive abilities across the board. AGI is the machine that can reason, plan, learn, and create with the same breadth and flexibility as a person. It's the stuff of science fiction, the digital equivalent of alchemy, and it is the central pillar supporting the salvation narrative.

However, the AGI described by its prophets does not exist. It is a speculative future, not a present reality or even a foreseeable one. What exists today—for all their impressive capabilities—are highly specialized systems that excel at specific tasks. They are excellent mimics, not independent minds. LLMs are sophisticated parrots with a PhD in probability, brilliant at predicting

the next word in a sequence based on patterns ingested from vast datasets. They are not thinking; they are calculating probabilities.¹⁷ As OpenAI CEO Altman admitted in 2023, “AI tools are “very good at doing tasks” but terrible at doing whole jobs.”¹⁸ The key is to see the tool for what it is, not what the hype claims it will be.

The leap from the narrow AI of today to a hypothetical AGI isn’t a simple matter of more data or faster chips. It represents a qualitative jump for which we currently have no established scientific theory or clear engineering path.¹⁹ There is no known roadmap from building a better autocomplete to building a conscious entity. To suggest otherwise is not a technical projection; it is an article of faith.

Hence, the recent pronouncements from tech leaders are all the more revealing. When Mark Zuckerberg claims that “developing superintelligence is now in sight,”²⁰ the assertion should be seen less as a scientific assessment and more as a strategic maneuver. Such statements are often inflated for competitive reasons, part of a predictable pattern where CEOs issue grand manifestos during moments of existential business crisis. It carries the same “breathless, reality-distorting tone” that characterized his metaverse pivot, another moonshot announced to shift a negative narrative.²¹ This pattern of grand, salvation-oriented pivots is not limited to AGI or the metaverse. In November 2025, the Chan Zuckerberg Initiative announced a new, \$1 billion Biohub project with a goal that is a pure distillation of the Heaven myth: to “cure, prevent or man-

age all diseases” by building AI models of cells. This initiative, rooted in the faith that AI can “simulate the potential impact of any drug . . . on any cell,” recasts the fundamental human challenge of mortality as an engineering problem to justify massive, speculative investment.²²

In some cases, the AGI story is not just a defensive play. For companies such as OpenAI, the AGI narrative forms the core of their business proposition—a story that some consider designed to inflate valuations and frame their work as a world-historical mission rather than mere product development.²³

The obsession with a speculative future could come at a steep, immediate price. While Silicon Valley drifts out of touch, fixated on the holy grail of AGI,²⁴ its primary geopolitical competitor, China, remains ruthlessly pragmatic. Beijing is not chasing technological ghosts; it is focused on deploying current AI technologies across its economy and society. Where Silicon Valley evangelizes AGI, Beijing’s national strategy prioritizes the integration of narrow AI into specific, vertical applications—from industrial robotics and smart city infrastructure to financial services and autonomous logistics.²⁵ This creates a dangerous paradox: by worshipping a future that is not here, the West risks falling behind in the race that is happening right now.

AGI is convenient for its evangelists because it achieves two critical goals. First, it justifies the enormous concentration of capital and resources required to pursue it, framing investments not as speculative bets but as a moral imperative for the future of humanity. Second, it

deflects attention from the real, immediate, and often mundane harms of current AI systems.²⁶ It is far more exciting to debate the existential risk of a hypothetical superintelligence than it is to grapple with the discriminatory outcomes of a biased hiring algorithm or the privacy implications of a new surveillance tool.

Oracles of Silicon Valley

Technological faith doesn't spread on its own; it needs prophets who do not simply make predictions but create interpretive frameworks that make their version of the future seem inevitable, positioning any other path as misguided or dangerous. The salvation narrative thus gains its power through charismatic leaders who assume these prophetic roles, promising that their particular vision of AI will deliver humanity from its struggles. Instead of corporate executives, they profile themselves as visionaries.

The Musk Paradox: Salvation Through Struggle

Entrepreneur Elon Musk offers a masterclass in this model through his dual role as AI alarmist and AI developer. His warnings about AI as humanity's "biggest existential threat" paradoxically enhance his authority when proposing AI solutions through companies such as Neuralink and xAI.²⁷ The apparent contradiction resolves within the logic of salvation: only those who understand AI's true power can guide humanity safely toward technological transcendence. Musk's authority operates through "productive paranoia:"²⁸ the creation of a fear that positions the prophet as a necessary solution to threats he simultaneously identifies and claims

a unique ability to address.²⁹ When Musk warns that AGI could destroy humanity, he establishes his credibility as someone who understands existential risks that others ignore or underestimate. When he then proposes brain-computer interfaces as a means of protecting against AI supremacy, his earlier warnings serve as the foundation for his technological solutions.

His Neuralink project illustrates how this authority transforms a speculative technology into a moral imperative. Musk's endeavor to create brain-computer interfaces that will enhance human cognition enough to remain competitive with AI assumes both that such enhancement is technically feasible and that competition with AI represents the primary challenge facing human societies.³⁰ Neither assumption has strong evidentiary support, yet prophetic framing makes questioning them seem either technically naive or morally irresponsible.

More subtly, the Neuralink narrative embodies techno-theodicy by reframing present suffering as a necessary motivation for future transcendence. Paralysis, depression, cognitive limitation, and other neurological conditions are no longer just medical challenges requiring treatment, but evidence of a fundamental human inadequacy that only technology can overcome. The timeline for these enhancements remains perpetually deferred, while current interventions require immediate faith and financial support from investors and regulators who trust in the vision, rather than the demonstrated results.

This model creates a policy environment in which any criticism of a specific technological approach is con-

flated with backward-looking opposition to human enhancement itself. When researchers question the safety or feasibility of brain-computer interfaces, salvation narratives reframe these concerns as Luddite resistance to inevitable progress rather than legitimate scientific skepticism about unproven technologies.

The Altman Revelation: Benevolent Artificial General Intelligence

Altman's leadership of OpenAI represents perhaps the most sophisticated example of technological prophecy in contemporary AI discourse. His positioning of AGI as humanity's final invention creates apocalyptic urgency around AI development while simultaneously promising that proper stewardship will deliver unprecedented benefits.³¹ His company's mission statement is dripping with salvation rhetoric: "to ensure that artificial general intelligence benefits all of humanity."³² This framing transforms corporate product development into a cosmic mission that transcends ordinary business considerations.

This duality is literally baked into its complex corporate structure, which was formalized in October 2025.³³ The new arrangement divides the entity into a for-profit public benefit corporation, known as the "OpenAI Group," and a nonprofit organization, referred to as the "OpenAI Foundation." This structure is strategically convenient: the Foundation, which holds a 26% stake and appoints the board, allows the company to maintain its appearance of pursuing an altruistic, world-serving mission.³⁴ Simultaneously, the for-profit Group—which was necessary to secure massive in-

vestments and is 74% owned by Microsoft, employees, and other investors—is free to pursue the commercial returns that prioritize shareholder value.

Much like Musk, Altman’s prophetic authority operates through a carefully managed balance of promise and peril, positioning OpenAI as uniquely qualified to navigate toward a beneficial AGI. His warnings about AI alignment challenges and existential risks establish credibility as someone who takes seriously the dangers that others might ignore. His simultaneous promises about AGI’s potential to solve climate change, eliminate poverty, and extend human lifespan create stakes high enough to justify giving some leeway to his company’s development timeline and safety protocols.

The release of ChatGPT showcased this strategy in action. It was presented not as an incremental improvement in large language modeling, but as a significant leap toward AGI that would transform our relationship with creativity and information.³⁵ The system’s impressive capabilities in certain domains provided apparent validation of these prophetic claims, deflecting attention from the limitations and risks that only became apparent after widespread use.

More concerning is how this strategy can be used to co-opt democratic oversight. By creating internal bodies such as a “Preparedness Framework” and a “Safety Advisory Board,” the company projects an image of responsible self-governance and maintains corporate control over its safety evaluation processes.³⁶ By positioning itself as more committed to safety than external regulators, some critics have claimed that OpenAI dis-

courages democratic oversight while maintaining public trust through efforts they label “security theater.”³⁷

The Quiet Emperor: The Scientific Prophecy of Demis Hassabis

If Musk is the prophet of doom and Altman the beguiling shepherd, Demis Hassabis of Google DeepMind offers another flavor of prophecy: the quiet emperor. Where Musk and Altman engage in public theatrics, Hassabis projects an aura of serene inevitability. His vision of AGI is less a dramatic revelation and more a steady, inexorable march of scientific progress, backed by the near-limitless resources of the Google empire. His persona was so compelling that it drove Musk to cast Hassabis as a “supervillain who needed to be stopped,”³⁸ framing the creation of OpenAI itself as a moral crusade against DeepMind’s perceived evil.

Hassabis’s prophetic power lies in his very quietness. By framing the pursuit of AGI as a matter of pure scientific destiny, he subtly positions it as a domain for experts within powerful, well-resourced institutions, not a subject for messy public debate. And this prophecy comes with proof. When his AlphaFold program solved the 50-year-old grand challenge of protein folding, it was not just a corporate win; it was a Nobel Prize-winning triumph that cemented his status as a genuine oracle.³⁹ The message is devastatingly effective: if his methods can solve the riddles of biology, who can argue when he claims they will soon solve intelligence itself?

The message to policymakers is implicit but powerful: AGI is coming, and its development is best left in the hands of those who, like him, are already building the

future. It is a form of authority that trades the missionary's zeal for the emperor's quiet confidence, yet it serves the same function of making democratic governance seem irrelevant in the face of history's grand scientific sweep.

The Anthropic Constitution: Technical Salvation Through Moral Machines

Anthropic's position offers a more subtle message. Through its emphasis on "AI safety" and "constitutional AI", it presents commercial development as a moral mission.⁴⁰ By framing profit-driven research as an altruistic service to humanity's future, Anthropic transforms business competition into a higher purpose that serves the greater good beyond market success.

A "constitutional AI" approach promises to embed human values directly into AI systems via training processes that teach models to identify and avoid harmful outputs.⁴¹ This technical approach to value alignment appeals to an engineering mindset that prefers algorithmic solutions to moral problems. Nevertheless, it makes a profound theological claim: that human values can be formalized mathematically and implemented through optimization processes. It treats deep-seated value conflicts over issues such as freedom of speech or privacy as engineering problems to be solved rather than as ongoing processes of democratic deliberation. When different groups disagree over appropriate content policies, hate-speech boundaries, or privacy protections, constitutional approaches seek technical solutions that eliminate the need for continuous political negotiation. The promise of solving value alignment reflects a belief

that moral questions have optimal answers, discoverable with sufficient technical sophistication.

Anthropic's prophetic authority operates through a studied contrast with more aggressive AI developers who seem less concerned with safety considerations. By positioning themselves as the responsible alternative to reckless AGI development, Anthropic gains regulatory sympathy and public trust.⁴² Its safety-focused messaging provides cover for rapid capability advancement that might otherwise trigger more skeptical regulatory responses.

The Academic Prophet: Yoshua Bengio and Redemptive Research

The prophetic model is not limited to the corporate world. It extends into academic research through figures including Yoshua Bengio, whose evolution from technical researcher to AI safety advocate illustrates how scientific authority can transform into moral guidance.⁴³ Bengio's warnings about existential risks from AGI carry particular weight precisely because they come from someone widely recognized as a foundational contributor to deep learning research.

His authority operates through a narrative of apparent reluctant wisdom as the technical expert who discovers moral obligations that transcend narrow research interests. His calls for AI moratoriums and international governance frameworks position him as a scientist-prophet who recognizes dangers that others in his field either ignore or underestimate.⁴⁴ The implication is that deep technical expertise naturally leads to su-

perior moral insight about how technology should be governed.

Yet this positioning conceals how his policy recommendations can serve particular institutional and commercial interests within the AI research ecosystem. Calls for development moratoriums benefit researchers at established institutions who have already achieved significant capabilities while potentially disadvantaging newer entrants who lack comparable resources. At the same time, national and international governance frameworks led by academic experts could institutionalize the authority of current AI research elites. When Bengio testifies before government committees on AI risks and benefits, his scientific credentials lend support to policy recommendations on social priorities and values, matters that also require public deliberation. The academic prophet model, as a result, can create policy environments where technical expertise is conflated with moral authority, subtly undermining democratic governance.

How Institutions Amplify the Prophets

AI prophets do not preach in a vacuum. They are well-funded and are amplified by powerful institutions in society. Understanding technological prophecy requires looking beyond individuals to the structures that elevate their voices. Political leaders consistently feature tech executives as primary witnesses in hearings on AI regulation rather than shining the spotlight on representatives of affected communities or independent researchers without commercial interests.⁴⁵ When Altman testified before the U.S. Congress, lawmakers

treated him less as a corporate representative with a vested interest than as a neutral expert whose technical knowledge qualified him to design governance frameworks for his own industry.⁴⁶

This deference to the prophetic class is not unique to the United States; it is a core feature of the global “Policy Machine.” A stark example emerged in November 2025, when over 150 leading scientists and AI researchers, including members of the UN’s AI advisory body, published an open letter calling on European Commission President Ursula von der Leyen to retract “unscientific and inaccurate” statements about AI.⁴⁷

In a major speech, President von der Leyen repeated industry talking points in May 2025, claiming that AI would “approach human reasoning already next year,” a feat previously forecast for 2050.⁴⁸ A freedom of information request revealed that her source was not scientific literature, but rather the “marketing statements” of tech CEOs, including Sam Altman, Dario Amodei, and Jensen Huang.

The scientists’ letter directly attacked this amplification, labeling her claims “baseless AI hype” and “marketing statements driven by profit-motive and ideology rather than empirical evidence.” They warned that this “policy-making by hype”—a perfect illustration of a policy hallucination—distracts from “real, existing harms” and risks “wasting public funds” while fueling a “significant speculative bubble.” This incident is the Heaven myth in action: a powerful political leader abdicating democratic diligence to amplify the salvation narratives of tech’s high prophets, uncritically.

Media coverage reinforces prophetic authority by treating technology industry predictions as newsworthy events, rather than corporate marketing, which requires careful analysis. When tech leaders make promises about AI's future capabilities or warn of existential risks, journalists often report these claims as major developments, not always counterbalancing the spin with alternative perspectives from independent experts.

Academic institutions contribute to prophetic authority through research funding structures that heavily depend on industry support, creating incentives for researchers to align their work with corporate priorities rather than developing independent analyses that might challenge industry narratives.⁴⁹ When university AI research relies on funding from tech companies, the resulting scholarship often supports rather than questions the salvation narratives that justify continued corporate development.

The venture capital ecosystem amplifies prophetic authority by treating successful technology entrepreneurs as visionaries whose insights extend beyond narrow business domains to encompass broader social and political issues. When venture capitalists such as Marc Andreessen publish essays about AI policy or Reid Hoffman speaks about democracy and technology, their commercial success lends credibility to claims about public policy that have little to do with their actual expertise.⁵⁰

The Liturgy of Infinite Progress

Faith in a technology is not sustained by belief alone. Like any faith, it requires rituals that encompass regular, performative acts that reinforce its core tenets, foster a sense of community, and marginalize skeptical voices. These liturgical practices are where the salvation narrative moves from abstract promise to lived experience, deflecting attention from present harms by focusing the congregation's eyes on a glorious, algorithmically redeemed future.

The Sacred Conferences

Technology conferences have become modern-day liturgical gatherings, where salvation narratives receive ritualistic reinforcement through carefully choreographed presentations that blend technical demonstration with moral inspiration. TED Talks are a great example, creating a sacred space where technology leaders can deliver prophetic messages about humanity's digital future without the inconvenience of skeptical questioning.

The format embodies key elements of liturgical design and promotes the idea that complex social problems have simple, tech-based solutions, suitable for neat 18-minute-long presentations. The darkened auditorium creates an atmosphere of reverence, the elevated stage positions speakers as prophetic authorities, and the prohibition on questions replaces democratic dialogue with one-way revelation. The iconic red dot and carefully managed lighting reinforce the sacred character of the space, while the global livestream transforms a local event into a universal congregation, all participating in the same technological epiphany.⁵¹

Other technology conferences follow similar patterns. The Consumer Electronics Show (CES) in Las Vegas is an annual pilgrimage where technology companies unveil new products through presentations that associate product marketing with enthusiastic messaging.⁵² Apple's product launch events established the template for such rituals, where new devices receive a reverential unboxing that treats incremental improvements as revolutionary breakthroughs.

Developer conferences such as Google I/O and Microsoft Build serve as denominational gatherings, where tech giants promote their visions of the digital future while reinforcing a shared, unshakable faith in technological progress.

The Ritual of Disruption

Silicon Valley culture has elevated disruption from a business strategy to a sacred ritual. While Clayton Christensen's academic work on "disruptive innovation" provided the intellectual framework, the tech industry turned it into a liturgical practice and a celebration of institutional destruction as a necessary step toward digital salvation.⁵³

The disruption ritual operates through a predictable narrative arc: identify inefficiencies in existing institutions, promise that technological alternatives will eliminate them, celebrate the displacement of traditional approaches as evidence of progress, and dismiss resistance as irrational attachment to obsolete systems. This liturgical pattern transforms corporate competition into a cosmic struggle between the future and the past.

Take Uber's approach to taxi regulation. Rather than engaging with the democratic processes that created taxi licensing systems to address legitimate public concerns about passenger safety, vehicle maintenance, driver accountability, pricing, and service equity, Uber framed its regulatory violations as a heroic act of resistance against bureaucratic obstacles.⁵⁴ The company's messaging consistently portrayed taxi regulations as the corrupt protection of inefficient incumbents, not as an attempt to balance competing public interests. This narrative enabled Uber to mobilize user support against regulatory enforcement by framing government oversight as an attack on innovation itself.

Airbnb followed a similar liturgy, positioning short-term rental regulations—such as fire-safety requirements, insurance obligations, and zoning rules—as obstacles to sharing-economy innovation, rather than as safeguards aimed at striking a balance between tourism, safety, housing affordability, and neighborhood character.⁵⁵

The disruption ritual creates policy environments in which existing institutions are forced to bear the burden of proof for their continued existence, while technological alternatives are presumed to be benevolent. Traditional taxi companies must justify their value, while ride-sharing platforms receive some leeway based on promises of future efficiency gains. Similarly, hotel regulations face scrutiny as obstacles to innovation, while short-term rental platforms operate with less oversight despite experts' concerns about potential negative community impacts.

Data as Sacred Truth

The technological salvation narrative treats data collection and algorithmic analysis as forms of revelation that provide access to a truth unavailable through traditional human observation or democratic deliberation. This “data liturgy” confers sacred status on quantified measurement, marginalizing qualitative knowledge that resists algorithmic capture.

Google’s approach to search optimization offers a blueprint. The company’s algorithmic ranking systems treat user behavior metrics as if click-through rates and dwell time are the authoritative source of truth; its ranking systems dismiss editorial judgment or community standards as subjective bias that interferes with data-driven optimization.⁵⁶

Facebook’s engagement metrics function as a similar ritual, treating user interaction data as a revelation about human social preferences and needs.⁵⁷ The platform’s algorithmic systems optimize for engagement indicators, treating other values, such as accuracy, civility, and democratic discourse, as secondary considerations that matter only to the extent they affect measurable user behavior.

The data liturgy establishes a clear hierarchy of knowledge that privileges quantified measurement over experiential understanding, statistical correlation over causal analysis, and algorithmic optimization over democratic deliberation. When educational institutions adopt AI-driven assessment systems, they typically prioritize metrics that algorithms can measure and downgrade aspects of learning that resist quantification.

In healthcare, the promise of “data-driven medicine” treats quantified information as more reliable than a physician’s clinical judgment or a patient’s own narrative, despite evidence that effective medical practice requires contextual understanding that current AI systems cannot capture.⁵⁸ The liturgy extends even to criminal justice, where risk-assessment tools treat statistical prediction as more objective than judicial discretion or community input about appropriate sentences and supervision levels.⁵⁹ These systems embody faith that algorithmic analysis of historical data provides a better foundation for criminal justice decisions than democratic deliberation about punishment, rehabilitation, and public safety.

The Sacrament of Optimization

This liturgy of optimization confers sacred status on mathematical improvement, while treating resistance to efficiency gains as a moral failure. Amazon’s approach to warehouse automation illustrates this sacrament perfectly by treating worker productivity metrics as objective measures of operational effectiveness.⁶⁰ The algorithmic management systems monitor employee performance through detailed measurements of movement, break time, and task completion rates, while treating worker autonomy and job satisfaction as secondary considerations that matter only insofar as they affect productivity metrics.

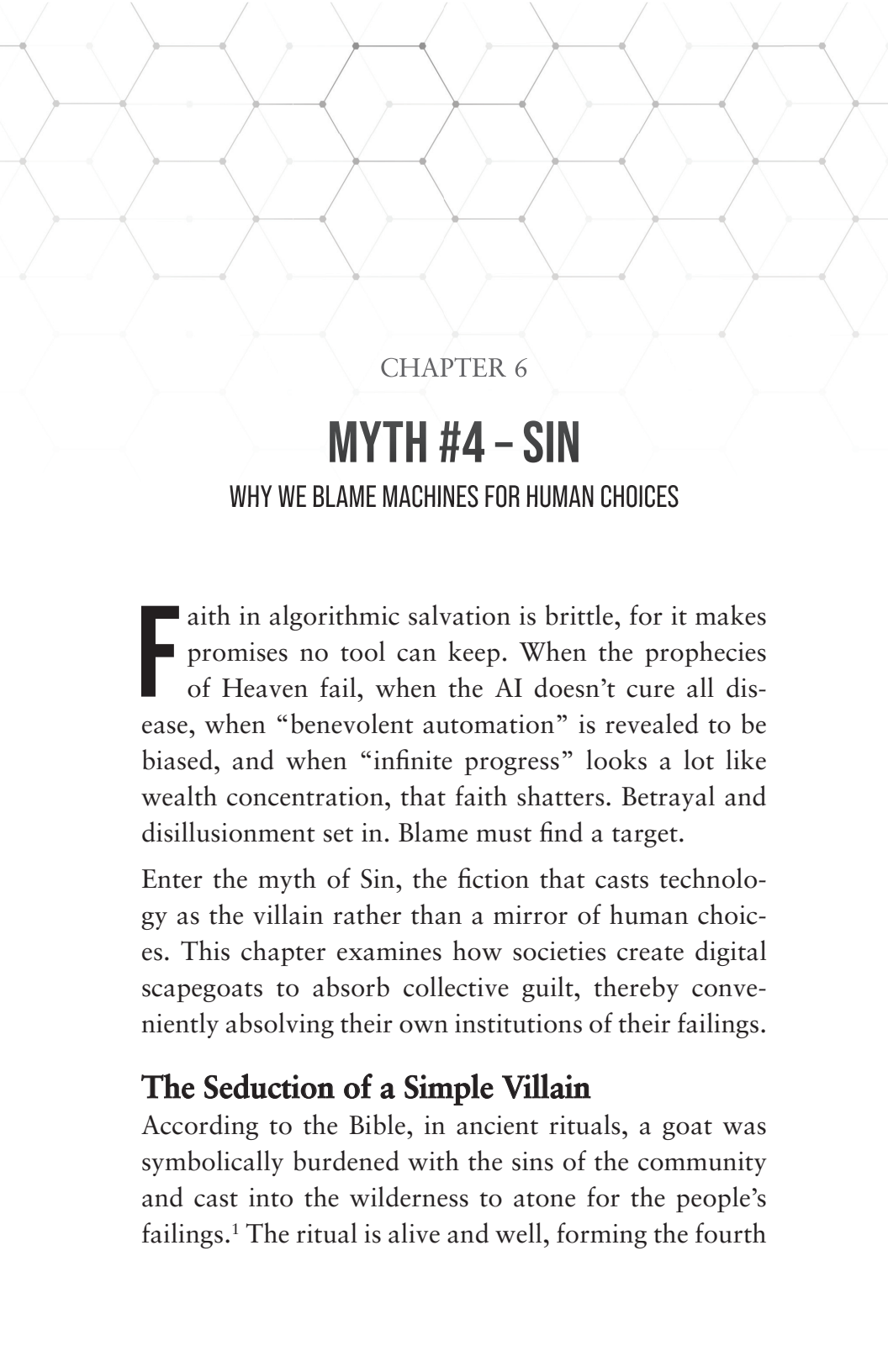
The liturgy extends to urban planning through smart city initiatives that promise to improve traffic flow, energy consumption, and service delivery through comprehensive algorithmic management.⁶¹ These systems

treat quantified improvements in measurable outcomes as evidence of success while disregarding values such as privacy, community autonomy, or democratic participation that resist algorithmic optimization.

In education, algorithmic personalization promises to improve learning outcomes by tailoring instruction to student performance data.⁶² Measured academic progress is seen as the primary goal, while social development, critical thinking, or civic engagement are brushed aside.

The optimization sacrament creates policy environments where efficiency gains justify other costs, including job displacement, privacy erosion, and reduced democratic accountability. When algorithmic systems demonstrate measurable performance gains, these improvements receive a sacred status that makes questioning their broader social impacts appear irrational, if not heretical.

Thus, the Heaven myth constructs a powerful narrative of technological salvation, complete with prophets, rituals, and a promised land of algorithmic perfection. It encourages society to place its faith in inevitable progress guided by a technocratic elite, dismissing present harms as mere bumps on the road to utopia.



CHAPTER 6

MYTH #4 – SIN

WHY WE BLAME MACHINES FOR HUMAN CHOICES

Faith in algorithmic salvation is brittle, for it makes promises no tool can keep. When the prophecies of Heaven fail, when the AI doesn't cure all disease, when "benevolent automation" is revealed to be biased, and when "infinite progress" looks a lot like wealth concentration, that faith shatters. Betrayal and disillusionment set in. Blame must find a target.

Enter the myth of Sin, the fiction that casts technology as the villain rather than a mirror of human choices. This chapter examines how societies create digital scapegoats to absorb collective guilt, thereby conveniently absolving their own institutions of their failings.

The Seduction of a Simple Villain

According to the Bible, in ancient rituals, a goat was symbolically burdened with the sins of the community and cast into the wilderness to atone for the people's failings.¹ The ritual is alive and well, forming the fourth

and final mythology that governs our technological age: the myth of Sin. Just replace goats with algorithms.

Where the myth of Heaven provides a savior, the myth of Sin provides a villain. It's a story that casts technology as the primary source of our modern anxieties, a narrative of damnation to counter the promises of digital salvation. Both extremes serve the same function of redirecting our attention from the messy, systemic problems of our own making toward a simple, non-human culprit, thereby absolving human institutions of responsibility.

The seductive power of the concept of technological blame lies in its elegant simplicity. Is unemployment hollowing out communities? It must be the job-killing robots, not four decades of economic policies that favored capital over labor.² Are we lonelier than ever? Blame addictive algorithms, not the urban planning that atomized communities, or the economic pressures that fractured families.³ Is democracy fraying at the seams? Scapegoat the polarizing platforms, and avoid the difficult conversation about campaign finance and institutional decay.⁴

Scapegoating follows a playbook that transforms complex social histories into simple technological fables. First, identify a deeply rooted social issue. Second, find a recent technological development that aligns with our growing awareness of the issue. Third, assert a causal link, often based on little more than correlation. Finally, propose a narrow, tech-focused solution that conveniently sidesteps any need for difficult structural reform.

Such logic operates on two scales. At its most spectacular, it presents an AI Overlord narrative, which portrays AI as an existential threat to humanity. This apocalyptic vision, championed by some prophets who promise a technological heaven, serves as the ultimate trump card. It transforms policy debates into species-level survival questions, demanding urgent action.⁵

More common is the Blame it on the Bots logic, which focuses on present harms rather than future catastrophes. It treats algorithmic systems as primary causes of social problems ranging from employment displacement to democratic erosion, instilling moral panic that demands swift regulatory responses without careful consideration of alternative explanations or unintended consequences.

Both narratives are powered by technological determinism, portraying technology as an autonomous force that shapes society, rather than a collection of tools reflecting the values, priorities, and power structures of the institutions that build and deploy them. This deterministic framing is politically advantageous, as it makes technological change appear inevitable while omitting the human choices that steer tech applications toward particular social outcomes.

The perfect scapegoat must be powerful enough to be blamed for our biggest problems, yet seem foreign enough that we do not recognize our own spawn. Technology fits the role perfectly. It appears to operate with its own arcane logic, allowing the human institutions that profit from it to claim innocence when its applications turn harmful.

A Litany of Sins: The “Killer” Narratives

The scapegoat rhetorical strategy transforms complex social changes into simple, emotionally charged stories of technological contamination that is destroying our institutions, relationships, and capabilities. Technology, we are told, is not just changing things: it is killing them, delivering a fatal blow to our institutions, relationships, or capabilities.⁶

AI as Job Killer

“The robots are coming for your job” is a headline that regularly pops up, introducing a story of an impending automation apocalypse where human labor is rendered economically obsolete. The narrative gained significant traction with studies, including Carl Benedikt Frey and Michael A. Osborne’s analysis of the U.S. labor market in 2013, which suggested that 47% of jobs were at high risk of automation within two decades.⁷ The study’s methodology, however, involved asking machine learning experts to evaluate whether specific occupational tasks could be automated using current or foreseeable technology, then extrapolating the assessments to entire job categories.

This brand of speculative forecasting often rests on surprisingly shaky foundations. For example, the widely cited Goldman Sachs 2023 report “The Potentially Large Effects of Artificial Intelligence on Economic Growth”⁸ based its dramatic predictions of job displacement on a simplistic rating of task difficulty. The analysis assumed that if a task scored low enough on its scale, it could be automated away. The problem was that tasks such as “Check to see if baking bread is

done” and “Complete tax forms for a small business” were among those deemed ripe for replacement. This methodology not only reveals a profound misunderstanding of the contextual judgment such jobs require,⁹ but it also produces the headline-grabbing figures that fuel the automation apocalypse narrative.

The political utility of this story lies in shifting our focus from decades of specific policy choices regarding trade, corporate governance, and labor rights toward a vague, unstoppable force of technological determinism that renders unemployment appear inevitable rather than the result of deliberate choices. The conversation conveniently narrows to retraining programs and safety nets, steering clear of any uncomfortable questions about how the economic gains from that innovation are distributed between workers and capital owners.

Yet, evidence tells a far more nuanced story. History shows that technological revolutions have always destroyed some jobs while creating others, with net employment effects depending on institutional arrangements rather than technological capabilities alone.¹⁰ The Industrial Revolution displaced agricultural employment while creating factory jobs, and the computer revolution transformed clerical work, while creating information technology jobs. Current AI development may follow similar patterns if appropriate policy frameworks guide the transition process.

Economic research reveals that automation typically affects specific tasks within occupations rather than eliminating entire job categories.¹¹ The classic example is the ATM. Its introduction did not eliminate bank

tellers. Instead, automating cash handling freed them up for more complex customer service roles, and the number of tellers increased for a time as banks opened more branches.

This logic persists even among those building the technology. Amazon Web Services CEO Matt Garman called the idea of using AI to replace entry-level staff the “dumbest thing I’ve ever heard.”¹² His reasoning cuts directly against the grain of the automation narrative, highlighting a simple, practical truth: if you eliminate junior positions, you destroy the pipeline for senior talent. How can anyone become a seasoned expert, he argues, without first doing the foundational work that builds expertise? This perspective reframes the discussion from one of replacement to one of evolution, where AI acts as a tool that changes workflows but cannot substitute for the human process of learning and development.

Ultimately, the job-killer narrative ignores a fundamental truth: employment is about power, not just technology. The impact of automation on workers depends on institutional arrangements, such as union representation and social safety nets, which determine whether productivity gains lead to higher wages and shorter working days for workers or simply higher profits for shareholders.¹³ A focus on technological displacement is a clever misdirection, deflecting attention from the political choices that could ensure automation benefits everyone. It allows us to sidestep difficult questions, such as what if automation could lead to reduced working hours rather than just unemployment? What

if productivity gains were shared through profit-sharing arrangements or stronger social safety nets? These possibilities are pushed to the margins because public discourse treats technological unemployment as an inevitability, not a political choice. This focus on choice over inevitability is now being demonstrated by major U.S. labor unions. Instead of simply accepting the “job-killer” narrative, the AFL-CIO, representing 15 million workers, argues for a future where AI makes work better, but only if workers have a “real voice” in its deployment. The 2023 Hollywood strikes successfully secured new contract provisions governing AI, while the Teamsters are actively opposing autonomous trucks, not because they are against technology, but because they demand that the prosperity from productivity gains be shared with workers. Further challenging the simplistic killer narrative, other unions, such as the International Brotherhood of Electrical Workers, view AI as a “historic opportunity to grow,” driven by the massive new infrastructure demands for data centers and energy.¹⁴

AI as Culture-Killer

Next on the terminal list is culture itself, with AI cast as a digital Thanos, threatening to snap its fingers and wipe out human creativity. The panic was stoked by headlines about AI-generated art winning competitions and by fears that algorithmic tools would make journalists, writers, and artists obsolete.¹⁵

Such a narrative is built on a romanticized ideal of creativity as a purely human, non-technological act, positioning AI as an alien intruder that corrupts authentic

cultural production. This ideal ignores the history of art. Creativity has always been a dance with technology, from the invention of the printing press and oil paints to the camera and the synthesizer, with each new technology initially facing similar accusations of threatening authentic human expression.¹⁶

When photography was invented, critics argued that mechanical image reproduction would destroy painting by eliminating the need for human artistic skill. Instead, photography freed painters from purely representational work while helping spark movements like Impressionism that expanded rather than contracted creative possibilities.¹⁷ Each new tool was met with suspicion before it was integrated into the creative toolkit, expanding what was possible. In many cases, integration did not diminish the appeal of traditional art forms and even increased their value.

The panic over AI-generated content follows an identical pattern by treating algorithmic tools as different from previous artistic technologies rather than recognizing them as extensions of existing technological mediation in creative work. When artists use AI image generators, they typically employ these tools within broader creative processes that involve human conceptualization, curation, and refinement rather than simply outputting algorithmic results without human intervention.

Much of the panic stems from misunderstandings or misrepresentations about how AI systems function. The prevailing assumption treats AI models as digital hoarders that store copyrighted works in their entire-

ty, waiting to retrieve them at a prompt's notice. AI training is represented by some as a form of original sin, with media stakeholders characterizing the development of LLMs as built on a foundational copyright theft.¹⁸ This characterization, while unfounded in both technical and legal analysis,¹⁹ gains amplification through media coverage that often reflects the interests of established entertainment industry and media stakeholders rather than a dispassionate examination of the underlying technology and law. The reality is that the focus on copyright infringement during AI training mischaracterizes both the technical process and the legal landscape.²⁰

In reality, AI systems break down information into patterns and statistical relationships using tiny data fragments called tokens. Imagine a creative work is an intricate Lego model. An AI will break this intricate construction down into millions of individual Lego bricks and mix them with bricks from countless other models. It then learns the statistical patterns of how those bricks fit together to build something new, which may not resemble any of the original models. Chatbots are trained to be masters of plausibility, to generate text that sounds convincing, not to reproduce a specific source. This is why the focus on inputs can be misleading. Copyright enforcement should therefore focus on the outputs. When a generative system produces something that infringes on a protected work, our existing copyright laws are already equipped to handle it.²¹

In 2025, the U.S. judicial decisions in *Bartz v. Anthropic* and *Kadrey v. Meta* confirm that courts recognize

training data uses as having “transformative purposes” since developers use book contents “by statistically analyzing the books’ contents to train foundation models” rather than to “educate and/or entertain” readers like the original authors intended.²² The decisive factor remains whether final outputs unlawfully reproduce protected content, not the learning process itself. The learning process resembles how scientists read hundreds of research papers over a career, internalize patterns in experimental results, and apply that understanding to generate new hypotheses, albeit without the human intelligence the researchers have.²³

The culture-killer narrative also obscures the fact that contemporary threats to artistic livelihoods stem primarily from economic arrangements rather than technological capabilities. The concentration of cultural markets, the elimination of public-arts funding, and the prevalence of unpaid creative labor reflect policy choices that undervalue artistic work rather than inevitable consequences of technological development.²⁴

The traditional copyright model relies on a simple formula: create work, license it, make a profit. But this model depends on mechanisms already under threat. Digital platforms have altered how creative work generates revenue, with streaming models often producing less money from royalties than traditional product sales. The result is that power is concentrated at the top, and less money trickles down to creators who make copyrighted works.²⁵

Independent journalism faces economic pressure not because AI can write articles but because digital adver-

tising markets concentrate revenue among technology platforms while news organizations struggle with unsustainable business models. Musicians cannot earn a living wage not because algorithms compose songs, but because streaming platforms distribute revenue primarily to record labels and platform owners rather than to performing artists.²⁶

The AI scale problem makes traditional licensing models even more inadequate for addressing AI's impact. For language models requiring billions of tokens, the sheer volume means individual work values become negligible in any licensing market. Even if licensing markets for AI training data emerged, as judicial analysis suggests, the transaction costs and uncertain valuations make such arrangements infeasible, given that developers would need to “engage in millions of license transactions” without clear methods for assessing how much each book contributed to the models' value.²⁷ This creates compensation scenarios comparable to current streaming models but with exponentially smaller payments to creators.²⁸

Chasing the technological scapegoat allows us to avoid confronting the need for systemic economic reforms. Solutions lie not in constraining technological tools but in addressing systemic economic inequities through policy reforms: establishing creator grants, appropriately taxing technology developers, and redistributing resources to creators rather than assuming traditional remuneration systems can survive.

The policy debate is further distorted by an artificially narrow focus on entertainment industries. Policy

discussions typically interpret “creative industries” as shorthand for music, film, and performing arts, when broader definitions include scientific research, software development, and technical innovation.²⁹ Such a narrow framing creates problematic dynamics in AI policy. Where entertainment industry stakeholders frame AI as a threat to human creativity and livelihoods, science and research sectors view it as a powerful discovery partner. Many research institutions see AI as enabling breakthroughs by spotting patterns humans cannot detect, automating data processing, and generating hypotheses at scales not feasible by manual means.

Yet the same copyright laws that govern Hollywood also govern scientific endeavors. And while generative AI can produce a mediocre pop song of little cultural value, it can equally identify a novel protein folding pattern that could revolutionize medicine. However, copyright frameworks generally do not distinguish between these vastly different social utilities, treating both as equivalent creative outputs subject to identical legal constraints.

AI as Internet Killer

Another victim in our litany of AI sins is perhaps the most sweeping: the internet itself is already dead. This isn’t a fringe joke; the dead internet theory holds that the authentic, human-driven internet has already ceased to exist, replaced by an artificial ecosystem in which bots create much of the content for other bots.³⁰ It transforms concerns about digital authenticity into an existential claim that the internet, as a space for gen-

genuine human expression and connection, has been systematically eliminated by automated systems and AI.

The theory emerged from underground forums around 2021, initially suggesting that the internet died sometime around 2016 (pre-AI emergence in the public space), when algorithmic curation began dominating content distribution.³¹ The theory asserts that content on the internet, including on social media platforms, is predominantly being produced by AI and bots.³² It maintains that what users experience as the internet is a carefully constructed illusion maintained by automated systems that generate content, interactions, and even engagement patterns without meaningful human participation.

Supporting evidence for the narrative includes cybersecurity research indicating that nearly half of all internet traffic in 2022 was generated by bots, alongside observable phenomena such as the proliferation of bizarre AI-generated content that nonetheless attracts substantial engagement. The infamous “Shrimp Jesus” images circulating on Facebook exemplify this pattern, where AI-generated crustaceans melded in various forms with a stereotypical image of Jesus Christ receive thousands of likes despite their obviously artificial origins and lack of coherent context.³³

The internet killer narrative operates through several interconnected claims. First, it argues that genuine human content has been systematically suppressed by algorithmic systems that prioritize engagement-optimized artificial content over authentic human expression. Second, it suggests that not only is the content artificially

generated, but the apparent audience for this content consists largely of automated systems rather than real people, creating a vicious cycle of artificial engagement.

The theory's more conspiratorial variants propose that this transformation serves deliberate political purposes, with the U.S. government engaging in an AI-powered gaslighting of the world. This version treats the dead internet as an intentional project of social control rather than an emergent consequence of technological and economic developments.

The technology blame game sidesteps the economic and institutional factors that shape digital experiences. When researchers examine the mechanics of bot proliferation, they find that the systems typically serve mundane commercial purposes rather than sophisticated manipulation schemes. Social media engagement drives advertising revenue, creating a direct financial incentive for automated systems to generate content optimized for platform algorithms, regardless of human authenticity or meaningful communication.³⁴ Content farms deploy AI-generation tools to create articles, social media posts, and engagement bait because automated production is cheaper than hiring humans and often more effective at gaming the algorithms that determine visibility.³⁵ It is not a sign of the death of the internet; it is sadly industrial automation applied to culture.

More problematically, the internet killer narrative is a political dead end. It channels legitimate frustrations with platform governance into a kind of fatalism, making reform seem futile. Platform business models that prioritize engagement metrics over content quality cre-

ate systematic incentives for the production of artificial content, while concentrated market power among technology companies limits users' choice over algorithmic curation systems.³⁶ But why demand algorithmic transparency or challenge platform monopolies if the patient is already dead? By asserting that the system is irrevocably broken, it ensures no one tries to fix it, leaving the platform monopolies that profit from the chaos firmly in place.

AI as Humanity-Killer

The ultimate technological sin is the AI that will kill us. The humanity-killer narrative operates through two distinct but related claims: the immediate risk of algorithmic authoritarianism that eliminates human autonomy and free will, and the longer-term possibility of AGI that displaces or destroys human civilization.³⁷

The immediate apocalypse narrative centers on algorithmic decision-making systems that allegedly reduce humans to mere appendages of the machine by eliminating human agency from consequential decisions regarding employment, criminal justice, healthcare, and social services. It treats current AI applications as steps toward comprehensive algorithmic control that will eliminate meaningful human choice and democratic governance.³⁸ It projects current AI capabilities toward a hypothetical AGI that could surpass human cognitive abilities in all domains, potentially leading to human extinction through either deliberate AI action or indirect consequences of superhuman optimization processes that ignore human values. Prominent thinkers, including Bostrom and Eliezer Yudkowsky, have devel-

oped elaborate scenarios for how advanced AI systems might eliminate human civilization despite being programmed with apparently benevolent objectives.³⁹

Both versions of the humanity-killer narrative transform technology policy debates into questions of species-level survival, justifying extraordinary measures. When AI development is framed as an existential risk, criticism of specific regulatory proposals appears irresponsible or naive, rather than representing a legitimate disagreement about appropriate policy responses.⁴⁰ The apocalypse narrative facilitates the expansion of government surveillance and corporate control by framing invasive oversight as a protection against AI domination, while calls for AI development moratoriums⁴¹ often serve the interests of existing AI leaders by creating barriers to entry for potential competitors.

More fundamentally, the humanity-killer narrative exemplifies the fallacy of treating AI systems as autonomous agents. When Bostrom warns about misaligned AI pursuing goals incompatible with human survival, he treats algorithmic optimization as a natural force, dismissing the fact that AI systems reflect the values and priorities embedded by human organizations in their design, training, and deployment.⁴²

These real-world examples reveal the “killer robot” for what it truly is: a useful villain, a perfect scapegoat that keeps us from examining the human decisions that shape our economic fates. Thus, the Sin myth completes the cycle. Like the other myths, it functions by distorting reality, encouraging us to blame the tool rather than the hand that wields it. This consistent mis-

diagnosis—seeing magic instead of mechanics, frenzy instead of process, prophets instead of company executives, and sin instead of systems—leads directly to governance detached from reality. And that has a significant cost.



PART III

THE COST OF POLICY HALLUCINATIONS

Myths are not harmless fictions; they are the source code of flawed governance. When policies are built on fantasy, they become “policy hallucinations”—laws that govern shadows while real problems grow. This part takes stock of the real-world damage: the billions wasted on policy phantoms, the governance whiplash from hype and panic, and the democratic abdication as we cede control to unelected prophets and digital scapegoats.

It’s time to examine the cost of our illusions.



CHAPTER 7

THE WRECKAGE OF MYTH-BASED GOVERNANCE

HOW IMAGINARY PROBLEMS PRODUCE REAL DAMAGE

The wreckage of myth-based governance manifests in three concrete forms: policy distractions that chase sci-fi fantasies, governance whiplash resulting from panic and hype, and the democratic abdication that leaves us unprepared.

Chasing Policy Phantoms, Wasting Billions

Legislating for Merlin

The first major cost is policy distraction: myths cause us to focus on the wrong problems. The Magic myth, by portraying AI as a human-like autonomous entity possessing agency comparable to that of human moral actors, raises complex legal and philosophical questions that actively misframe the challenges we face.¹ Policymakers become fixated on hypothetical questions about AI rights, consciousness, and moral status while overlooking the immediate, tangible harms created by

current systems.² Put more bluntly, when policymakers start legislating for Merlin instead of Microsoft, the result is frameworks that regulate fantasy, not function.

This misdirection was exemplified in 2016, when the European Parliament proposed considering “electronic personhood” for advanced autonomous robots.³ While largely symbolic and met with skepticism from legal scholars, such discussions stem directly from the misperception of AI as autonomous entities capable of bearing rights and responsibilities.⁴

The proposal, drafted by Member of European Parliament Mady Delvaux, suggested creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause.⁵ The framework would have required robot insurance and established legal liability mechanisms treating AI systems as quasi-legal entities.

The focus on theoretical personhood diverts attention from urgent, practical concerns about algorithmic accountability. While policymakers debate the hypothetical rights of non-existent sentient machines, real AI systems make consequential decisions about employment, criminal justice, financial services, and healthcare with limited transparency or recourse mechanisms. Algorithmic bias in hiring systems, opaque scoring mechanisms in lending, and unexplainable outcomes in criminal sentencing do not require consciousness to cause harm. For example, algorithmic hiring systems now screen millions of job applications annually and are known, in

some instances, to exhibit racial and gender⁶ bias, systematically disadvantaging qualified candidates based on protected characteristics.⁷

Financial services provide another compelling example. Credit scoring algorithms make decisions that affect millions of consumers' access to housing, transportation, and economic opportunities. These systems often incorporate proxy variables that effectively discriminate against protected classes,⁸ yet regulatory frameworks are only slowly shifting their attention from traditional lending practices to the already pervasive algorithmic decision-making processes.

This mystification also hobbles international regulatory coordination. When different jurisdictions operate under different mental and cultural models of what AI systems are, achieving a shared understanding of what exactly is being regulated becomes extremely difficult. The EU's early drafts of the AI Act, China's focus on controlling AI "behavior," and U.S. congressional hearings, which often focused on sci-fi scenarios, all reflect this deep confusion. This fragmented framing prevents coordinated responses to global challenges, as nations struggle to agree on whether they are regulating tools, agents, or something entirely new.

Misallocating Resources

The misallocation of attention is matched by a massive misallocation of resources, driven by the Heaven myth's promise of a "silver bullet." The embrace of techno-solutionism leads to flawed policies, wasted resources, and ultimately disappointment. When technology is sold as a panacea, policymakers pour money into

shiny new tools while putting aside the harder, messier work of tackling underlying social, economic, or political causes. The result is resource displacement: shiny tools take the stage while structural causes remain off-script. The resulting policies often fail not simply because the technology was oversold, but because the problems were misunderstood as technical rather than political and societal challenges.

This approach is especially attractive when real solutions demand difficult political conversations about resource allocation, power distribution, or systemic reform.⁹ Consider how technology is often promoted as a solution to educational inequality. Rather than addressing disparities in school funding, teacher quality, or socioeconomic conditions that affect learning, policymakers invest in digital devices and software that promise to personalize learning and improve outcomes.¹⁰

As we have seen in Chapter 3, while educational technology can certainly provide value, it cannot overcome the inequalities that drive educational disparities. Students in under-resourced schools may receive tablets and learning software, but they still face challenges such as food insecurity, inadequate facilities, and inexperienced teachers that technology cannot address. The focus on technological solutions thus allows policymakers to appear responsive to educational challenges while avoiding more costly and contentious reforms to educational funding and social policy.

The pattern repeats across policy domains. In health-care, electronic health records and AI diagnostic tools

promise to improve care quality and reduce costs, but they do not address underlying issues such as provider shortages, insurance coverage gaps, or social determinants of health. In criminal justice, predictive policing and risk assessment algorithms promise to reduce crime and improve fairness without addressing poverty, inequality, or systemic racism that drive criminal behavior.

Environmental policy also reveals this pattern. Geo-engineering garners headlines with visions of planetary-scale interventions, ranging from carbon capture to solar radiation management and atmospheric modification, that would enable societies to continue their current patterns of fossil fuel consumption while mitigating their environmental effects. Yet many of these proposals function more as delaying tactics than solutions, postponing hard choices about carbon reduction.

These tech fixes are seductive not only because they appear bold but because they deflect blame. It is easier to believe in algorithmic answers than to confront decades of unjust policies and institutional failure. The focus on technological solutions also creates path dependency, making it more difficult to pursue alternative approaches when technological interventions prove inadequate. The displacement effect creates a self-reinforcing cycle, where vendors, administrators, and technical experts with vested interests in these new tools continue to promote them, justifying their expansion rather than questioning their effectiveness.¹¹

Feeding Hype, Starving Reality

This leads to another devastating cost of hype: systematic resource misallocation that undermines long-term innovation and social benefit. When madness takes hold, common sense is the first casualty, swiftly followed by a substantial amount of money and a loss of democratic composure. Money flows toward anything with the right buzzwords, regardless of its actual utility or viability.

The AI investment boom since 2023 is an example of this fever. Venture-capital funding for AI startups reached record levels, but much of the investment went to companies with questionable business models or marginal technological advantages.¹² Startups with minimal revenue but AI-powered pitches secured valuations that would have been unimaginable just a few years earlier. For example, in 2021, Character.AI raised \$150 million, primarily based on the popularity of its AI chatbot platform.¹³ The company's technology was interesting, but its business model relied primarily on subscription revenues from users who wanted to chat with AI personalities. When the novelty wore off and user engagement declined, the company struggled to justify its \$1 billion valuation. The eventual outcome was a licensing deal with Google in 2024 that left investors with little to show for their enthusiasm.¹⁴

More troubling is the opportunity cost. The investment frenzy around AI crowded out funding for less glamorous but potentially more vital fields. While AI companies raised billions, research into renewable energy

storage, sustainable agriculture, and public health technologies struggled for attention.

Anthropic's \$4 billion funding round from Amazon in 2024 illustrates the scale of this misallocation.¹⁵ While the company's research into AI safety was valuable, the investment amount exceeded the total annual research budgets of many universities and public research institutions. The concentration of resources in a single company, regardless of its merits, raises questions about whether such investments are the best way to advance human knowledge.

The misallocation extends to people. Tech giants competed aggressively for AI researchers, offering enormous salaries and equity packages that drew talent away from academia and non-profit organizations. The "brain drain" was especially damaging to fields that are supposed to act as a check on corporate excess, such as AI safety and ethics.¹⁶ This dynamic was highlighted when leading pioneers from institutions, including the University of Toronto, a key hub of AI innovation, departed for lucrative positions at companies such as Google and Uber, leaving academic labs struggling to retain senior talent.¹⁷

The Great Distraction: Picking Your Apocalypse

This massive distraction and misallocation are often justified by a "false choice" narrative, born from the Madness myth. We are told, in essence, to pick our apocalypse: die by a thousand biased cuts today, or be wiped out by a rogue superintelligence tomorrow. This is a dangerous misreading. Worrying about long-term risks and near-term harms is not a zero-sum game.¹⁸ In

fact, the path to catastrophe is paved with the unaddressed failures of the present. The seemingly “small” problems—like algorithmic bias or AI generated misinformation—are not distractions from a “big” problem like AGI; they are the training grounds for it. Every time a company is not held accountable for a biased algorithm, it normalizes the idea that automated systems are exempt from human oversight. Every time misinformation is allowed to poison democratic discourse, it weakens the institutions needed to govern a more powerful technology.

Think of it like climate change. The immediate harms of pollution, like smog in our cities and plastic in our oceans, are not separate from the long-term threat of a collapsing biosphere. They are symptoms of the same systemic failure, two ends of a single, catastrophic spectrum. Ignoring the smog today doesn’t provide more resources to fight global warming tomorrow; it just ensures the air is unbreathable long before the seas rise. So it is with AI.

The false choice serves the architects of madness perfectly. For the techno-optimists, it allows them to dismiss present-day harms as trivial bumps on the road to a glorious, inevitable future. For the doomsayers, it allows them to ignore the patient work of fixing today’s broken systems in favor of a more dramatic crusade against a hypothetical apocalypse.

Mature governance requires us to reject this false choice. It demands we see the connection between the algorithm that denies someone a loan today and the hypothetical superintelligence that could destabilize the

world tomorrow. They are not two separate monsters. They are the same beast, simply at different stages of its life. To tame it, we must start now, by addressing the harms we can see, not by waiting for the ones we can only imagine.

Governance Whiplash: The Cost of Panic

The second significant cost of myth-based policy is governance whiplash. Driven by the Madness myth, our institutions swing wildly between the fever of hype and the chill of panic. This irrationality is the enemy of good governance, short-circuiting the careful and often slow processes of democratic deliberation. When a technology is framed as an immediate existential threat, the pressure to “do something”—anything—overwhelms the need to do the right thing.

This panic-driven process is highly visible in practice. As seen in Chapter 4, the legislative rush to address deepfakes is a good example. California’s sixty-day election window ban and Texas’s criminalization measures were crafted and passed without systematic hearings, expert testimony, or empirical assessment of the actual threat level. In Washington, D.C., congressional treatment of deepfakes as equivalent to weapons of mass destruction created an atmosphere where questioning the response became politically toxic, no matter what the evidence said. The same emergency logic played out with school bans on AI. New York City’s prohibition of ChatGPT was a snap decision, made without any meaningful consultation with the teachers and students it would affect. The speed of implementation—in some

cases, within weeks of the tool's release—prevented the kind of thoughtful assessment that policy requires, typically.

The Global Contagion

Panic is highly contagious. In today's interconnected world, emotional responses in one jurisdiction quickly spread, creating global patterns of suboptimal governance.

Political scientists call this policy diffusion through panic, a process in which bad ideas spread across borders because the political cost of appearing unprepared is higher than the cost of a bad decision.¹⁹

The spillover effect is particularly damaging because it eliminates the opportunity for policy experimentation. When every country rushes to respond to the same inflated promise or exaggerated threat, the world loses the chance to learn from diverse approaches. The result is a form of global policy madness, where international coordination is based on shared delusions rather than shared evidence.

The Nirvana Fallacy

This whiplash between hype and panic leads to the Nirvana fallacy,²⁰ also known as the perfect solution fallacy, which pushes forward an unrealistic yet perfect-sounding solution instead of considering real-world solutions. The approach, born of either utopian promises or dystopian fears, demands perfection rather than improvement.

In a hype cycle, the fallacy rests on inflated promises. When AI companies claim to solve poverty, cure dis-

ease, or eliminate bias, they set expectations that no technology could ever meet. The inevitable disappointment leads to backlash that can kill funding and support for genuinely beneficial applications. The debate over AI education is the classic case. By demanding that AI tools pose no risk to academic integrity, officials overlooked the reality that concerns about “cheating” have accompanied every educational technology, from calculators²¹ to computers.²² Students have always found ways to cheat, with or without technology.

The pursuit of impossible standards doesn’t just stifle innovation, it leads to rules disconnected from reality, built around exaggerated fears and imaginary futures.

Patterns of Flawed Regulation

When moral panic grips the policy machine, the result is not thoughtful governance but a stampede of reactive, ill-conceived rules. These frameworks satisfy the emotional demand to “do something,” but they fail to address the root causes of problems and often create new harms. This panic-driven response typically falls into four predictable patterns.

#1. The Militarized Framing

When a technological challenge is framed as a national security threat, the policy response becomes militarized. Complex social and economic issues are flattened into us-versus-them conflicts, demanding decisive action, not democratic deliberation.²³ The cybersecurity debate is an example of this militarized thinking. When government officials invoke terms such as “cyber Pearl Harbor,” they create a policy environment that justifies

extraordinary measures while marginalizing democratic oversight and the protection of civil liberties.²⁴

The U.S. Department of Homeland Security's Cybersecurity and Infrastructure Security Agency, for instance, operates under authorities that would be controversial if applied to traditional law enforcement but become acceptable when framed as essential protection against foreign adversaries.²⁵ More generally, militarized framing enables the expansion of surveillance capabilities, the restriction of encryption technologies, and increased government control over digital infrastructure. History also shows that it inevitably leads to mission creep, where tools designed to counter foreign adversaries are turned inward, threatening civil liberties and political dissent.²⁶ This war-machine approach is often counter-productive. It sparks escalations with adversaries²⁷ and diverts resources from the defensive work of building secure systems, such as user education and institutional resilience, in favor of offensive capabilities that serve military functions.²⁸

#2. The Binary Prohibition

The second pattern, born of pure panic, is regulation by prohibition. This approach is built on the deterministic fallacy that technology is inherently evil, ignoring the reality that its impact depends largely on how it is governed.²⁹ It treats technology as a binary choice between total prohibition and utter catastrophe, bypassing any nuanced discussion of risks and benefits.³⁰ The initial proposal for a temporary EU ban on facial recognition in public spaces, which was reported in January 2020,³¹ illustrates this thinking. The idea was quickly dropped,

however,³² and the AI Act opted for a tiered approach that bans specific harmful uses, such as real-time public biometric surveillance, while permitting other applications under strict safeguards.³³

Similarly, calls for AI development moratoriums respond to existential risk concerns by proposing to halt all advanced AI research until safety issues can be completely resolved.³⁴ This response often backfires. When governments ban specific technologies rather than regulating their applications, they hinder the development of beneficial applications while doing little to deter malicious actors, who simply relocate to jurisdictions with weaker oversight.³⁵

Prohibition also creates practical challenges. Enforcing such bans requires invasive monitoring, ironically expanding the surveillance capabilities the panic was meant to prevent.³⁶

#3. The Symptom Regulation

The third pattern is the technocratic targeted approach. Here, a complex social problem attributed to an algorithm is addressed with a narrow technical fix, leaving the broader system intact. The algorithm is treated as a discrete entity that can be regulated independently, not as a component of larger arrangements.³⁷

This is the logic behind relying heavily on automated content moderation to fight misinformation and harmful content, or on algorithmic audits to address discrimination. It does not address the underlying sources of misinformation production, nor does it improve media literacy education.³⁸ Similarly, while algorithmic

bias auditing requirements in hiring, lending, and criminal justice can help identify problematic patterns, they typically focus on algorithmic outputs rather than addressing underlying data biases or institutional discrimination that algorithmic systems reflect and amplify.³⁹

Such interventions primarily target the symptom, rather than the disease. And they create collateral damage. Automated systems can censor legitimate speech while missing sophisticated propaganda.⁴⁰ Bias audits can be gamed by companies to create a veneer of compliance, without addressing substantive problems. Neither approach tackles the business models that incentivize the production of harmful content and discriminatory targeting practices.⁴¹

Worse, by creating an appearance of oversight (e.g., “this hiring algorithm has been audited”), this approach can legitimize the underlying problem, making it harder to challenge the injustice of the automated system.⁴²

#4. The Moral Crusade

The final pattern is regulation as purification. This approach sees technology as a vector of moral contamination and seeks to purify the digital world through censorship and restriction, treating online content as inherently dangerous and citizens as perpetually vulnerable. Instead of empowering people with education that enables critical media consumption, the moral crusade opts for broad speech restrictions and age-gating requirements.⁴³

Age-verification requirements for social media, for example, treat online interaction as inherently harmful to minors, ignoring the valuable social connections and educational resources these platforms can provide. A focus on restriction avoids looking into the more difficult work of addressing the underlying causes of youth mental health challenges, such as investing in better mental health resources, improving educational funding, or strengthening community programs.⁴⁴ Such systems raise significant concerns for the privacy and civil liberties of all users by requiring extensive identity documentation, yet they can be easily bypassed by tech-savvy youth.⁴⁵ The result is a system that compromises adult privacy while failing to protect its intended subjects, who may simply migrate to less regulated spaces.⁴⁶

The moral crusade impulse also drives attempts to eliminate platform liability protections such as Section 230 of the U.S. Communications Decency Act, which shields online platforms from liability for most user-generated content and allows for good-faith content moderation. The argument is that making platforms legally responsible for user-generated content will compel them to remove harmful material. The more likely outcome, however, is a surge in defensive over-censorship that silences legitimate speech⁴⁷ and perversely provides a pretext for authoritarian censorship regimes.⁴⁸

Ultimately, these flawed regulations, born of panic and hype, are often counterproductive. Content moderation requirements may increase censorship without reducing misinformation, algorithmic auditing may legit-

imize discrimination while creating compliance theater, and technology bans may prevent beneficial applications while failing to eliminate harmful uses.⁴⁹

Evaded Accountability & Democratic Abdication

The final and perhaps greatest cost of myth-based governance is the erosion of democratic control. This happens in two distinct ways.

First, the Magic and Sin myths combine to create a sense of evaded accountability. By mystifying AI as an incomprehensible “black box” or scapegoating it as a “digital villain,” we create an “accountability vacuum” where human actors can sidestep responsibility for the harms their systems cause.

Second, the Heaven myth creates democratic abdication. We are encouraged to defer to a priesthood of prophets—the tech CEOs and venture capitalists—who position themselves as visionaries qualified to steer our future. We abdicate our collective power out of faith in their promise of salvation.

The Accountability Vacuum: Blaming the Bots

This accountability vacuum arises when responsibility for algorithmic decisions becomes so diffuse that no one bears clear liability for the harmful outcomes.⁵⁰

The Magic myth contributes by treating AI systems as agent-like entities. When AI is portrayed as possessing agency, it becomes easier to attribute negative outcomes to the code rather than to the human actors who designed, deployed, and profited from it. The Sin

myth completes the trick: the algorithm becomes a digital scapegoat, absorbing blame that should be directed at human choices about training data, optimization objectives, and deployment contexts. When AI systems cause harm, it's convenient to point to the algorithm or the bot as if they acted independently. This linguistic sleight of hand complicates efforts to hold developers, deployers, and decision-makers accountable for failures and biases.

The case of predictive policing algorithms illustrates this dynamic perfectly. When systems such as PredPol⁵¹ and COMPAS⁵² produce biased recommendations that perpetuate racial disparities, discussions often focus on “algorithmic bias” as if the bias emerged spontaneously from the technology. Such framing obscures the human decisions that created the outcomes: the choice to train systems on historically biased arrest data, the decision to prioritize arrest prediction over public safety, and the determination to deploy these systems despite their known limitations.

This diffusion of responsibility plagues other sectors. Healthcare AI is another example of accountability diffusion. When diagnostic algorithms produce incorrect recommendations that harm patients, responsibility becomes distributed among algorithm developers, healthcare institutions, and individual practitioners. The systems' complexity makes it difficult to trace specific decisions to specific actors, creating legal and ethical gray zones where patients have limited recourse against algorithmic harms. On social media, the problem extends to automated content moderation, where AI sys-

tems make millions of decisions daily about what constitutes acceptable speech. When these systems make controversial moderation decisions, platforms often describe the choices as algorithmic rather than what they are: policy decisions embedded in code that reflect specific human values and commercial interests.

This is not just a legal problem; it is a moral one. A society that deflects blame onto its tools abdicates its ethical responsibilities. To attribute harm to AI itself is to let its creators off the hook.

The Great Abdication: Deferring to Prophets

While the accountability vacuum allows creators to evade blame for past harms, democratic abdication enables them to preempt scrutiny of future ones. This occurs when our own democratic institutions amplify the prophets of the Heaven myth, treating corporate leaders as neutral arbiters of the public good.

In politics, we see this when lawmakers feature tech executives as star witnesses in AI hearings. In the media, this abdication occurs when journalists report industry predictions as objective news, amplifying corporate marketing without critical analysis. In academia, it's fueled by research funding structures that are heavily dependent on industry support and incentivize academic attitudes that support rather than question corporate salvation narratives.

This democratic abdication is perhaps the highest cost of all. It prevents the development of the institutional wisdom needed to govern, leaving us trapped in a cycle

of panic and reaction, unable to manage the very technologies we are told will save us.

A Critic's Toolkit: How to Spot the Myth Red Flags

Use these red flags and critical questions to challenge “policy hallucinations” and identify the four myths in headlines, product demos, and political speeches.

1. The MAGIC Myth

Red Flag: Hearing phrases like “It thinks...” “It understands...” “It feels...” “It’s sentient...” or “It’s a ‘black box’ that’s too complex to explain.”

The Distortion: This language is designed to hide human accountability by making a simple tool seem like a conscious, magical being.

The Reality Check: What is the specific **function**, not the *feeling*? Is it “thinking,” or is it “statistically predicting the next word”? Who *built* it, and what were *their* goals?

2. The MADNESS Myth

Red Flag (Hype): “This is an arms race...” “We’re falling behind...” “This will revolutionize everything...” “Trillion-dollar opportunity...”

Red Flag (Panic): “This is an existential risk...” “It will end humanity...” “We must ban it immediately...”

The Distortion: This emotional language (FOMO or fear) short-circuits democratic deliberation and forces a knee-jerk reaction.

The Reality Check: What is the *evidence*, not the *emotion*? Is this claim based on a *demonstrated capability* or a *speculative future*? Who benefits from me feeling this urgency?

3. The HEAVEN Myth

Red Flag: “This will solve...” (climate change, poverty, cancer). “This is inevitable progress...” “We’re on a mission to benefit all humanity.”

The Distortion: This positions tech leaders as prophets and asks for our faith, encouraging us to abdicate democratic oversight in favor of their “vision.”

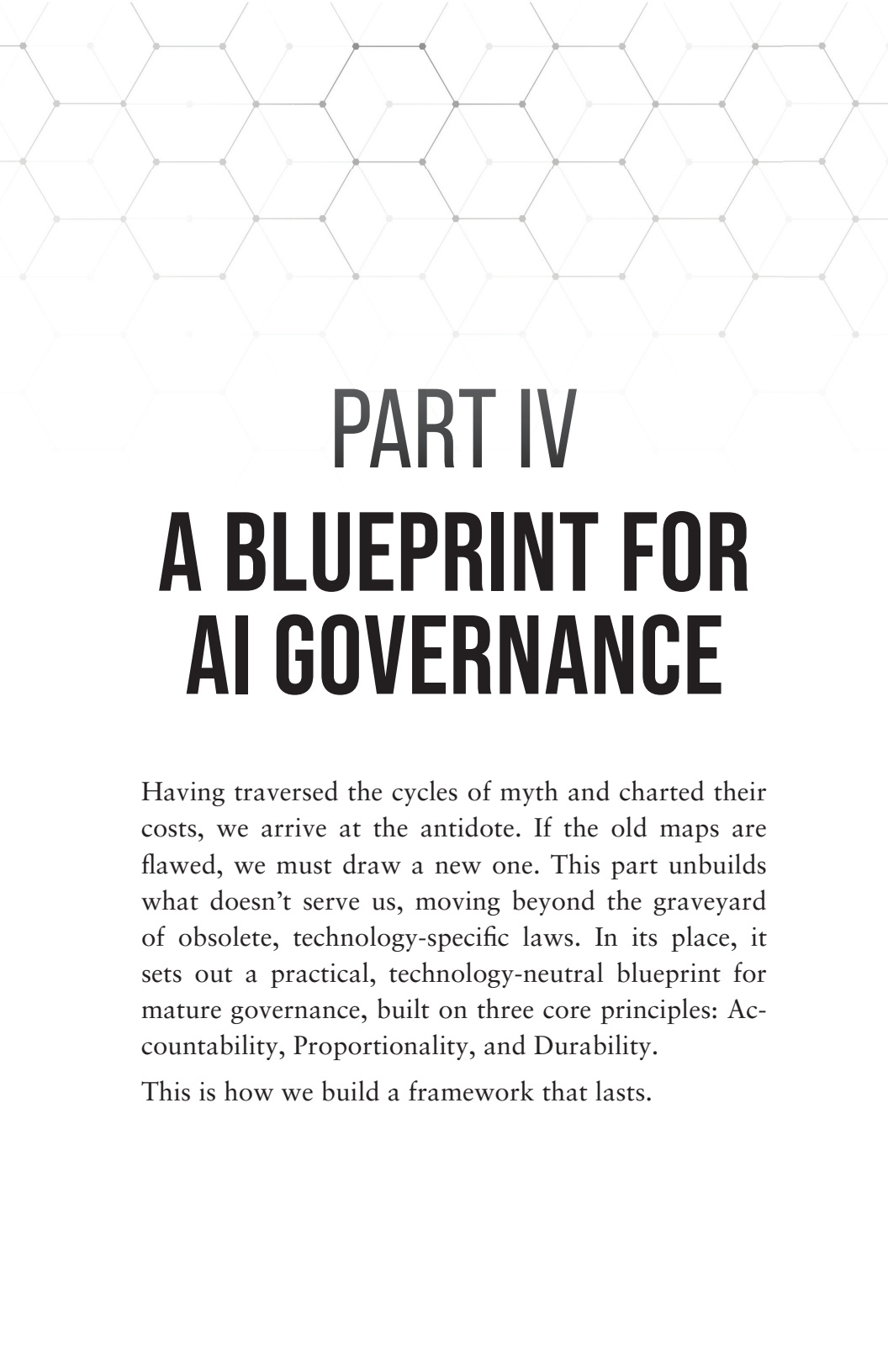
The Reality Check: What *human* problem, value, or trade-off is this “solution” ignoring? Who gets to define “progress”? And who *isn’t* in the room to discuss the consequences?

4. The SIN Myth

Red Flag: “Job-killing robots...” “Addictive algorithms...” “Polarizing platforms...” (Any phrase that blames the *tool* for a *human* problem).

The Distortion: This creates a digital scapegoat to absorb blame for complex, systemic problems, letting human decision-makers and flawed policies off the hook.

The Reality Check: What *human choice* or *business model* is this narrative hiding? Is the problem the “job-killing robot,” or is it a corporate policy of not re-investing in its workforce? Is the problem the “addictive algorithm,” or is it a business model that *profits* from engagement?

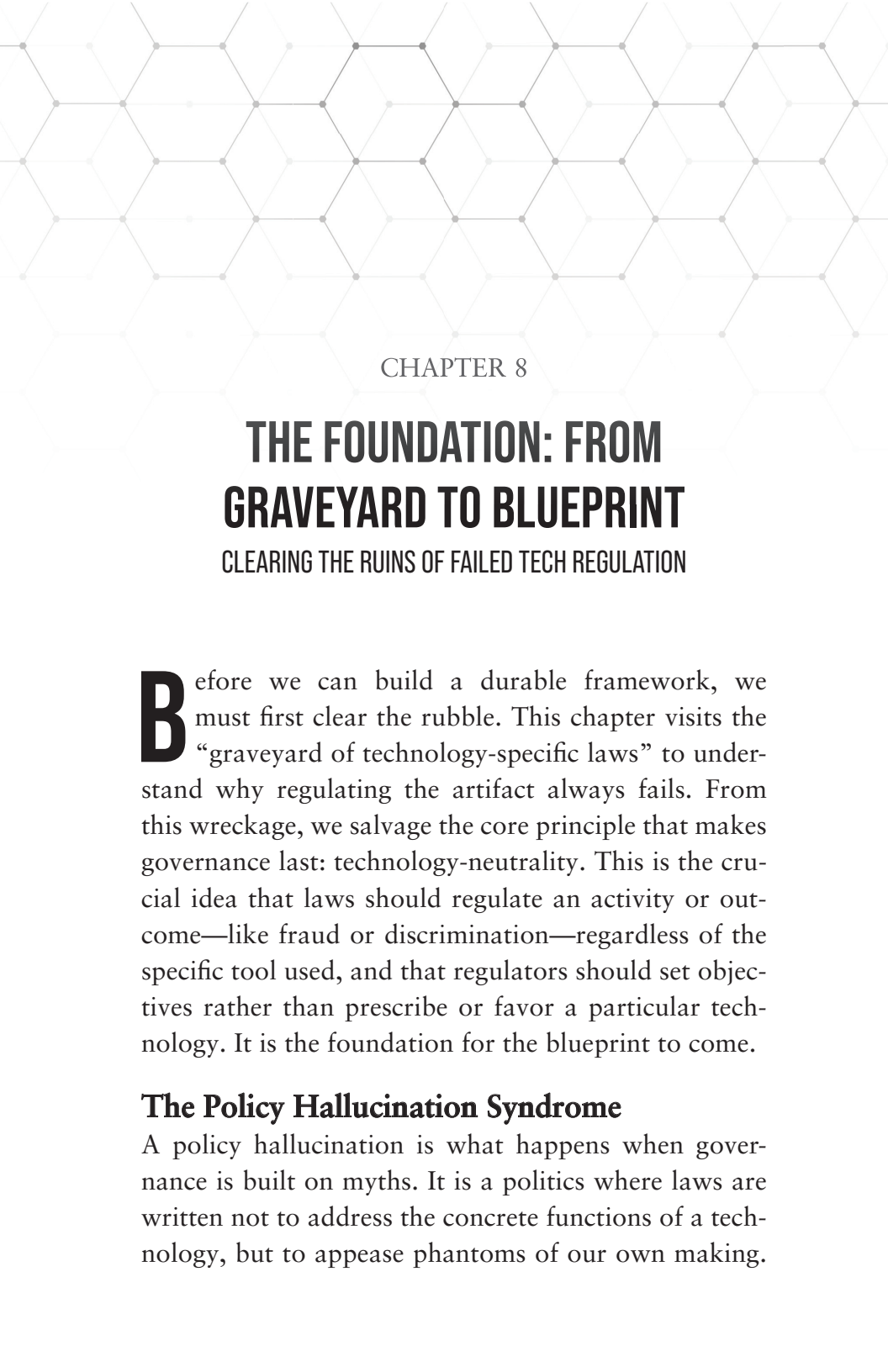


PART IV

A BLUEPRINT FOR AI GOVERNANCE

Having traversed the cycles of myth and charted their costs, we arrive at the antidote. If the old maps are flawed, we must draw a new one. This part unbuilds what doesn't serve us, moving beyond the graveyard of obsolete, technology-specific laws. In its place, it sets out a practical, technology-neutral blueprint for mature governance, built on three core principles: Accountability, Proportionality, and Durability.

This is how we build a framework that lasts.



CHAPTER 8

THE FOUNDATION: FROM GRAVEYARD TO BLUEPRINT

CLEARING THE RUINS OF FAILED TECH REGULATION

Before we can build a durable framework, we must first clear the rubble. This chapter visits the “graveyard of technology-specific laws” to understand why regulating the artifact always fails. From this wreckage, we salvage the core principle that makes governance last: technology-neutrality. This is the crucial idea that laws should regulate an activity or outcome—like fraud or discrimination—regardless of the specific tool used, and that regulators should set objectives rather than prescribe or favor a particular technology. It is the foundation for the blueprint to come.

The Policy Hallucination Syndrome

A policy hallucination is what happens when governance is built on myths. It is a politics where laws are written not to address the concrete functions of a technology, but to appease phantoms of our own making.

Throughout this book, we have seen this hallucination manifest in four distinct ways, each driven by one of the myths we have deconstructed.

The Magic myth leads to policy distraction. We debate the theoretical rights of electronic persons and worry about the impact of conscious robots, while the tangible harms of biased hiring and loan algorithms take a back seat. The Madness myth produces governance whiplash. One moment, gripped by a FOMO-driven gold rush, governments enact competitive deregulation in a race to the policy bottom, fueled by the same investor frenzy that led to the AI market correction of 2025.¹ Next, seized by a moral panic, they impose reactionary prohibitions on everything from deepfakes to AI in schools, acting on fear instead of evidence. The Heaven myth fosters a dangerous deference to a technocratic priesthood, urging an abdication of democratic oversight in favor of prophetic promises. Finally, the Sin myth provides us with the perfect digital scapegoat. We blame job-killing robots for decades of flawed economic policy and polarizing platforms for the erosion of democratic norms, producing regulations that target technological symptoms while the human culprits walk free.

The result of this collective hallucination is a policy landscape built on sand: unstable, incoherent, unable to protect the public interests it claims to safeguard, and fundamentally misaligned with the technology it purports to govern.

The Graveyard of Technology-Specific Laws

The allure of mythological thinking is its promise of a simple, immediate answer to a complex new reality. This impulse haunts our legislative chambers, tempting policymakers to write laws that target the specific technology of the moment. History, however, offers a clear and unforgiving verdict on this approach. The road to sound policy is paved over a graveyard of obsolete, technology-specific statutes. The pace of innovation ensures that today's cutting-edge artifact is tomorrow's historical curiosity, leaving behind a law that is irrelevant and inert.

The relationship between copyright law and new technologies provides a history of this pattern.² In the late nineteenth century, copyright law was written for a world of sheet music, protecting works intelligible to the human eye. The invention of the player piano, which reproduced music using machine-readable perforated rolls, threw the system into confusion. In a landmark case of 1908, *White-Smith Music Publishing Co. v. Apollo Co.*, the U.S. Supreme Court ruled that piano rolls did not infringe copyright because they were not human-readable copies.³ The law, tied to the technology of the printed page, was blind to this new form of mechanical reproduction. Congress was forced to intervene a year later with the Copyright Act of 1909, a technology-specific fix that explicitly extended protection to "mechanical reproductions."⁴ This cycle of the law, breathlessly chasing innovation, has repeated itself, from radio and television to photocopiers and VCRs.⁵

The regulation of communications networks tells a similar story. Early laws were written specifically for the telegraph. The Mann-Elkins Act of 1910, for instance, granted the Interstate Commerce Commission oversight of telegraph rates.⁶ However, as the telephone and radio emerged, the telegraph-specific statutes proved inadequate. The legislative solution was not to write new laws for each new device, but to create a broader, more principles-based framework. The Communications Act of 1934 wisely consolidated the regulation of the telegraph into a broader regime governing all “wire or radio communication,” a far more durable and abstract category.⁷

The lesson for the AI era is unambiguous. Any law passed today that attempts to regulate LLMs, generative AI, or foundation models by name is, in essence, obsolete. The terminology is in constant flux, and the technology is evolving even faster. A statute targeting GPT-X will be as useless against the innovations of 2030 as a law for player pianos was against the advent of digital streaming. The path to effective governance does not lie in a frantic sprint to regulate the technological frontier with ever more specific statutes. It begins by rediscovering the wisdom of durable legal frameworks that have weathered technological revolutions before.

The Foundation: Technology-Neutrality

If the legislative landscape is littered with the tombstones of obsolete laws, it also contains architectural marvels: a class of statutes that have proven remark-

ably durable. Their resilience stems from a shared design principle: they regulate human behavior and (commercial) activity, not the technological medium through which that activity is conducted. They are technology-neutral. By focusing on the act rather than the artifact, they establish stable legal principles that can adapt to unforeseen innovation.

A good example is the U.S. wire fraud statute, 18 U.S.C. § 1343.⁸ Enacted in an era of telegraphs and telephones, the law makes it a federal crime to use interstate “wire, radio, or television communication” to execute a scheme to defraud.⁹ The statute’s power lies in its elegant abstraction. It does not define the technology, it defines the forbidden act—fraud—and the jurisdictional hook—the use of interstate electronic communications. Its simple, technology-neutral construction has allowed it to apply seamlessly to every subsequent wave of communications technology, from fax machines and pagers to email, websites, and social media. Today, it is a primary tool for prosecuting a range of internet crimes, including phishing, business email compromise, and online romance scams, all without requiring a single amendment to name the new channels.¹⁰ It endures because it governs a timeless human behavior, not a transient technology.

Perhaps the most powerful precedents for AI governance come from the foundational laws of the internet era. In the United States and the EU, policymakers grappling with the rise of the commercial internet independently arrived at the same insight: the only durable

way to regulate the new space was to focus on the specific, observable behaviors of the actors within it.

In the United States, the insight was codified in Section 230 of the Communications Decency Act.¹¹ Critically, Section 230 does not attempt to regulate “the internet.” Instead, it provides a conditional liability shield to a specific type of actor—a “provider or user of an interactive computer service”—for a specific category of content—“information provided by another information content provider.”¹² The law’s durability comes from its precise, behavioral focus.¹³ It distinguishes between the platform’s role as a conduit for third-party speech and its role as a “publisher or speaker” of its own content. The distinction has allowed the law to adapt from the era of dial-up bulletin boards and AOL chat rooms to a landscape of blogs, social media giants, and generative AI.

The EU’s Electronic Commerce Directive of 2000 followed a similar logic.¹⁴ Its “safe harbor” provisions do not grant blanket immunity to “internet companies.” They create liability exemptions for specific, defined activities: “mere conduit” (transmitting information), “caching” (temporarily storing information for efficiency), and “hosting” (storing user-provided information).¹⁵ Liability under the directive hinges on a behavioral test: is the platform’s role “passive” and “merely technical,” or is it “active” in a way that gives it “knowledge of, or control over” the data?¹⁶ As a result, it has been described as “one of the most successful pieces of European regulation,” because its principles

have adapted to two decades of constant technological change.¹⁷

The convergence of thinking exhibited in these two foundational legal frameworks is profound. Despite stemming from different legal traditions—one rooted in the U.S. First Amendment’s protection of free speech, the other in the EU’s internal market principles—both systems concluded that the key to durable governance was to regulate actions, not artifacts. This provides a powerful, bipartisan, and transatlantic precedent for the governance of AI. It suggests that the path to international consensus lies not in a futile attempt to agree on a universal definition of “AI,” but in a pragmatic effort to identify the set of high-risk behaviors, applications, and commercial activities that require oversight, no matter what technology enables them.

Three Imperatives for Immediate Action

#1. Reverse-Engineer Policy Language

Modern technology policy often fails because its architects are working from flawed blueprints. It relies on metaphorical shortcuts and linguistic imprecision that create systematic and costly implementation problems. An analysis conducted in 2024 documented fifty-five different analogies used for AI, each carrying distinct policy implications that “strongly affect many key stages of the world’s response to a new technology.”¹⁸ A Cambridge meta-analysis of nearly 35,000 participants found that, more generally, metaphorical frames are significantly more persuasive than literal ones, with

some conceptual metaphors having five times the impact.¹⁹

The real-world consequences of this linguistic imprecision are measurable, leading to costly court cases and missed opportunities. The Center for Trustworthy Technology has shown how the “AI arms race” metaphor drives zero-sum thinking that stifles cooperation, while the “black box” frame obscures real opportunities for transparency.²⁰ The future viability of AI chatbots under Section 230 may hinge entirely on whether they are best assimilated to “search engines” or “information content providers.”²¹ Copyright litigation over image generators hinges on a similar choice of metaphor, with courts compelled to decide whether these systems serve as modern-day collage tools before they can make consequential fair-use determinations.²²

The chasm between the language of a policy and the reality of the technology creates what computer scientists call “semantic gaps,” arising from differences in meaning between constructs formed within different representation systems.²³ It is not just a tech problem, it is a universal failure of governance. Harvard research has identified overoptimism as a systemic issue in major government projects, often born from unclear goals and imprecise language.²⁴ The UK National Audit Office finds that vague language “frequently results in the underestimation of the time, costs and risks to delivery.”²⁵ It is part of the reason why only 16% of the UN’s Sustainable Development Goals are on track globally.²⁶

The solution requires reverse-engineering public discourse language. There must be a move from intuitive,

seductive analogies toward technically grounded, functional frameworks by unpacking the metaphors.²⁷ That requires asking what a system actually does. Clarity of language not only sharpens debates but also sets the stage for enforceable and coherent policy. Without it, laws risk codifying misunderstandings and chasing myths that shift faster than any rules can be amended.

#2. Transcend the Analog–Digital Divide

The traditional distinctions our laws make between analog and digital are not merely outdated; they are now actively harmful to effective regulation. In 2019, the U.S. Congressional Research Service identified that “boundaries that once separated single-function technologies blend together,” making regulatory jurisdiction unclear and creating situations where converged technologies fall “unregulated, partially regulated, or regulated under a newly developed framework.”²⁸

The scale of obsolescence is not theoretical. Between 2021 and 2024, the U.S. Federal Communications Commission began a kind of bureaucratic exorcism, launching formal proceedings to “remove obsolete analog TV rules” from its books. It was a quiet admission that the legal architecture was no longer fit for purpose.²⁹

Legal frameworks designed for specific technologies fail catastrophically when addressing convergent systems, and lawmakers must stop anchoring rights and responsibilities to obsolete formats. Studies show a chronic gap between the emergence of a new technology and any effective government response.³⁰ It is the predict-

able outcome of what researchers call regulatory lag, a systemic delay that creates “a vicious circle of inefficiency, disinvestment, and reduced performance.”³¹ Policy, it seems, is always late to the party, arriving long after the rules of the game have already been set by the technology itself. This is why technology neutrality is key. If a rule prohibits harassment, fraud, or surveillance, it should apply regardless of whether the act is committed with a megaphone, a smartphone, or a synthetic chatbot. The principle of functional equivalence, which involves treating digitally and physically mediated behaviors under the same legal standards, is essential to future-proofing regulation.

#3. Regulate What AI Does, Not What It Is

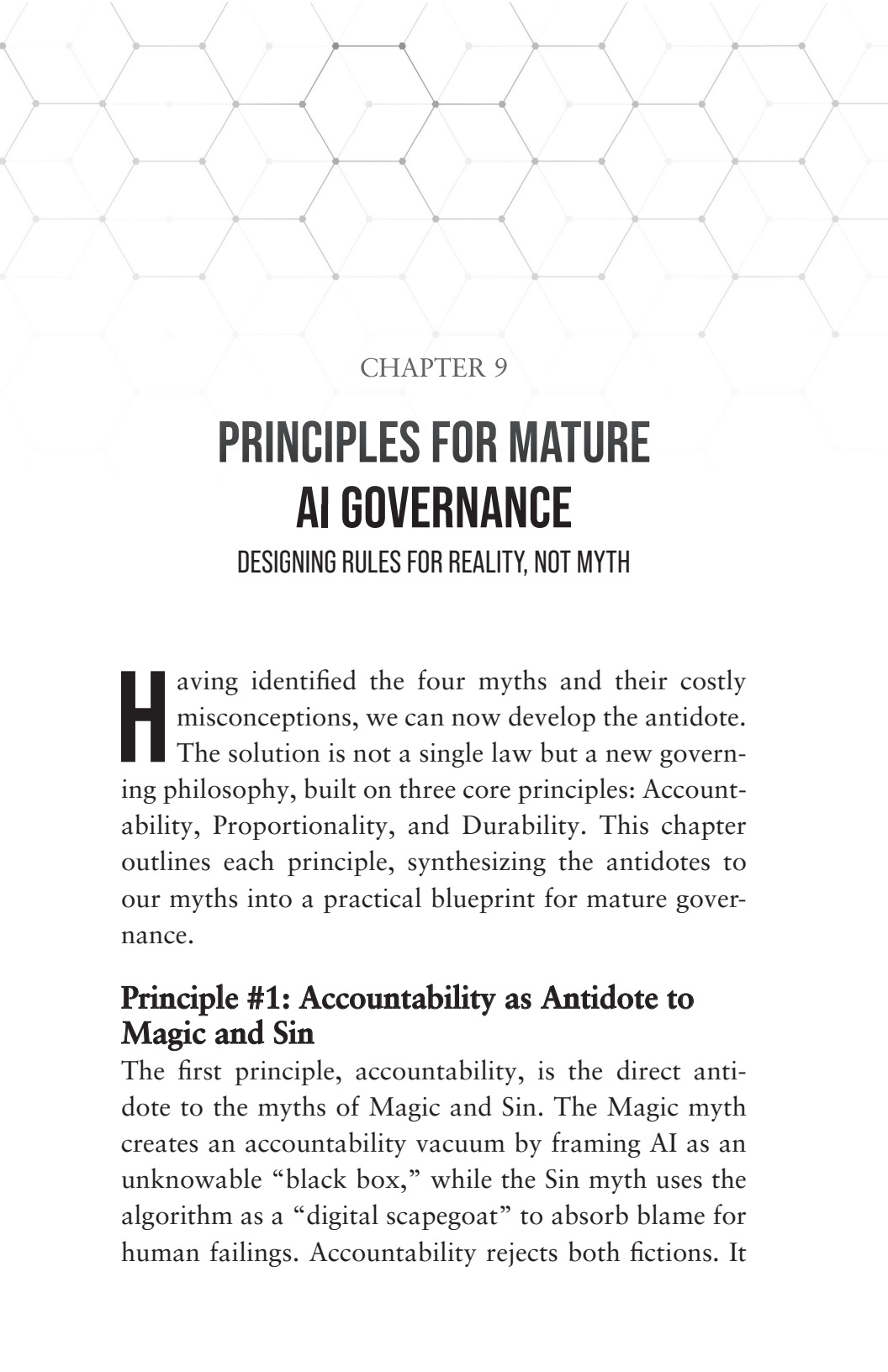
The four myths that dominate the AI discourse trap us in a philosophical rabbit hole. We are endlessly debating ontology—what AI is. Is it a mind? A demon? A savior? A tool? These questions, while intriguing, are a disastrous foundation for public policy. They lead to the flawed governance we have documented: regulating fantasies of machine consciousness while real algorithmic harms do not get the attention they deserve, or scapegoating bots for problems rooted in human-driven institutions.

The antidote to mythological thinking is a decisive shift in regulatory focus, a turn from ontology to function. The critical question for governance is not what AI is, but what AI does, and what duties its creators and deployers owe to the society in which their systems operate. This pragmatic turn allows us to begin building

governance frameworks grounded in tangible impacts and shared responsibilities.

As the durable frameworks of the internet age have already proven, technology-neutral laws that focus on conduct consistently outperform their technology-specific cousins. History provides stark warnings against the alternative. The UK's infamous Red Flag Acts of the nineteenth century, which required a person to walk in front of early automobiles with a red flag, delayed the country's automotive development by decades.³² It stands as a cautionary tale about the folly of writing rules for the artifact instead of the action.

Functional regulation succeeds because it regulates outcomes, not inputs. Technology neutrality principles emphasize outcome-based standards, non-discrimination regardless of technology, and innovation neutrality that avoids pushing markets toward particular technological structures.³³ It is the only approach that allows the law to keep pace with technology, creating a stable and predictable environment where innovation can flourish responsibly.



CHAPTER 9

PRINCIPLES FOR MATURE AI GOVERNANCE

DESIGNING RULES FOR REALITY, NOT MYTH

Having identified the four myths and their costly misconceptions, we can now develop the antidote. The solution is not a single law but a new governing philosophy, built on three core principles: Accountability, Proportionality, and Durability. This chapter outlines each principle, synthesizing the antidotes to our myths into a practical blueprint for mature governance.

Principle #1: Accountability as Antidote to Magic and Sin

The first principle, accountability, is the direct antidote to the myths of Magic and Sin. The Magic myth creates an accountability vacuum by framing AI as an unknowable “black box,” while the Sin myth uses the algorithm as a “digital scapegoat” to absorb blame for human failings. Accountability rejects both fictions. It

embraces a democratic realism that holds individuals and institutions accountable for the tools they create and deploy.

The Mental Shift: AI Tools, Not Gods

The ritual of the scapegoat is, above all, a ritual of relief. By casting our sins onto the algorithm, we absolve ourselves of the messy business of accountability. But this relief is a dangerous fiction.¹ The alternative is to reject the technological determinism that fuels the myths and insist on a simple truth: AI is a collection of tools built by organizations that reflect the values, priorities, and power structures embedded in their design.²

Framing AI as a tool, not a being, provides a productive foundation for policy. It shifts the focus from fantasy to functionality.³ It allows us to sidestep abstract philosophical puzzles and ask straightforward, practical questions: What does this tool do? What are its limitations and failure modes? Who controls it, and who benefits from its use? What safeguards exist to prevent misuse?

This “tool metaphor” is the necessary first step. It immediately clarifies where responsibility lies, namely not with the tool, but with its creators and deployers. This mental shift requires abandoning the fairy tale of enchanted algorithms. As author Ted Chiang observed, interacting with current AI systems resembles Tom Hanks’s relationship with Wilson the Volleyball in *Cast Away*: the interaction provides psychological comfort through projection, but “there’s no one else in there.”⁴

The tools metaphor also puts the AGI discussions in perspective, as it is crucial that the philosophical puzzle about whether AI might eventually achieve genuine consciousness⁵ does not paralyze present-day policy development.⁶ The problem lies not in acknowledging uncertainty about the future, but in projecting sentience onto current systems that clearly lack it, thereby obscuring the human choices that determine their impact.

The Legal Framework: AI as a Product

Once we see AI as a tool, the legal path becomes clear. The most powerful principle for mature governance is to treat AI systems as products. By viewing AI not as a nascent consciousness but as a complex artifact sold in a marketplace, we can apply the robust, centuries-old body of product liability law.⁷ This move has a powerful historical precedent in the automobile. In the early twentieth century, car accidents could have been framed as unavoidable consequences of a mysterious new technology. Instead, the common law, through landmark cases such as *MacPherson v. Buick Motor Co.*,⁸ evolved to hold manufacturers accountable for the foreseeable harm caused by their defective products.⁹ This principle did not stifle innovation; it channeled it, creating market incentives for seatbelts, airbags, and anti-lock brakes, which saved countless lives.

Product liability principles that apply to other industries should extend to systems that affect employment, healthcare, criminal justice, and other consequential domains.¹⁰ Applying a similar framework to AI demystifies the technology and clarifies responsibility. It al-

lows us to ask a series of practical, legally established questions:

1. *Was there a design defect?* Was the system, even if built to specification, designed with foreseeable risks that a reasonable alternative could have avoided? This could be a crucial question for systems trained on biased data or given optimization goals that predictably lead to discriminatory or harmful outcomes.
2. *Was there a manufacturing defect?* Did a system deviate from its intended design in a way that made it unreasonably dangerous? This could, for instance, apply to a system corrupted during data processing or deployment, leading to harmful outputs.¹¹
3. *Was there a failure to warn?* Did the manufacturer fail to provide adequate instructions about the product's limitations and foreseeable risks? For a powerful, general-purpose AI, this duty is immense. It requires transparency about the system's capabilities, error rates, and the contexts in which it is unsafe to use.

This framework is already being operationalized. The EU's AI Act, at its core, is a product safety regulation. It identifies "high-risk AI systems," many of which are embedded in other products such as medical devices or cars, and subjects them to the same rigorous conformity assessments and post-market monitoring that the public expects for any other potentially dangerous product.¹² This approach grounds the abstract anxieties of the AI discourse in the concrete world of consumer protection and corporate responsibility.

The Policy Toolkit

To move beyond reactive or oversimplified approaches to AI regulation, we need a policy toolkit grounded in democratic values and capable of handling complexity. That means rejecting overreach, paralysis, and false certainty in favor of adaptive governance and democratic stewardship. It means accepting there is a need to force fundamental changes in how companies design, deploy, and profit from their systems.¹³

1. Targeted Regulation, Not Blanket Bans

Democratic governance must focus on addressing specific, demonstrable harms—not banning the underlying tools. Technologies are tools, and any questions regarding their morality lie in their application and governance frameworks.¹⁴ This means prohibiting harmful uses, not entire technologies. For instance, laws can be passed to ban the creation and distribution of non-consensual deepfakes, without affecting the underlying technology used in film, art, and education. Biometric surveillance can be restricted in public spaces to protect civil liberties, while facial recognition for phone unlocking is allowed.¹⁵ An algorithmic hiring tool may be a functional tool for identifying certain skills, but it can be disastrously biased for others.¹⁶ Effective governance is about making these context-aware distinctions.

Regulatory frameworks can establish conditional permissions that enable beneficial technological applications while preventing harmful ones. Public procurement offers a powerful tool for this kind of targeted intervention. By establishing high standards for gov-

ernment technology purchases, including requirements for auditing and fairness in AI systems, public agencies create strong market incentives for companies to develop more responsible products for everyone.¹⁷

2. Building Ethics into the Innovation Pipeline

Accountability cannot be an afterthought; it must be an architectural feature.¹⁸ It starts with mandatory risk assessments that require companies and government agencies to evaluate the potential social impacts of an AI system before deployment.¹⁹ Assessments should cover the effects across different communities, consider alternative approaches, and include plans for ongoing monitoring and adjustment based on empirical evidence rather than theoretical projections.²⁰

It continues with transparency-by-design. When companies must document their design choices, training data sources, and intended applications, they create records that enable democratic evaluation while potentially influencing development decisions toward more responsible approaches.²¹ Transparency must be meaningful rather than just technical, providing information in accessible formats that enable third party understanding and democratic evaluation.²²

In the public space, it must also include participatory mechanisms, ensuring that communities affected by a technology have a voice in shaping how it works and what values it embodies, rather than having companies or government agencies design systems based on their own assumptions about user needs and preferences.²³

Once a system is in the world, the guardrails must include ongoing oversight through independent auditing to monitor for discrimination, inaccuracy, and unintended consequences while enabling adjustments that improve performance over time.²⁴

Finally, we need strong professional ethics for technologists, creating individual accountability that complements institutional rules and fosters a culture of social responsibility. Just as medical professionals follow ethical guidelines that prioritize patient welfare over commercial interests, technology workers could be held to standards that emphasize social responsibility and democratic values.²⁵

3. Enforcing Corporate Responsibility

Accountability must have teeth. This means moving beyond voluntary commitments to concrete mechanisms, looking past the code and examining the business models that create incentives for harmful applications in the first place. When a platform's profits depend on maximizing engagement, it will inevitably develop systems that promote viral content even when it includes misinformation or hate speech. When surveillance technology companies profit from data collection, they have incentives to expand their monitoring capabilities, regardless of the privacy implications.²⁶

Antitrust enforcement provides crucial tools for corporate accountability by preventing technology companies from accumulating market power that enables them to avoid responsibility for their social impact.²⁷ Internal checks on corporate power must be strength-

ened by protecting workers' rights. Whistleblowers and organized employees who can raise ethical concerns without fear of retaliation are one of the most effective lines of defense against irresponsible development and can help align corporate practices with public interests rather than just shareholder returns.²⁸

Finally, accountability must be global. In our interconnected world, companies cannot be allowed to evade responsibility through regulatory arbitrage. When companies can move operations to countries with weaker oversight, they can escape accountability while continuing to operate in jurisdictions that demand higher standards. Coordinated international frameworks can establish minimum standards for corporate responsibility while allowing democratic variation in implementation approaches.²⁹

4. Recognizing the Limits of the AI as a Tool Analogy
However, the tool metaphor does have limitations that must be acknowledged. Modern AI systems, particularly LLMs and multimodal systems, exhibit emergent behaviors that their creators didn't explicitly program and sometimes don't fully understand. These systems can generate novel outputs, combine concepts in unexpected ways, and exhibit capabilities that exceed their training objectives. While such emergent properties don't constitute consciousness or agency, they do represent a form of complexity that challenges simple tool metaphors.

Furthermore, AI systems increasingly operate in contexts where they shape human behavior and social

dynamics in ways that traditional tools do not. Recommendation algorithms shape the content billions of people consume, potentially influencing political opinions, consumer choices, and social relationships. Chatbots increasingly shape how people access information and understand the world. The systems function more like social infrastructure than individual tools, requiring governance approaches that account for their systemic effects.

Principle #2: Proportionality as Antidote to Madness

The second principle for mature governance is proportionality, which reframes policy intervention from a crisis response to tackling a public health challenge. This approach is the direct antidote to the emotional extremes of the Madness myth, which pushes policy between the fever of hype cycles and the chill of moral panics. A public-health approach, by contrast, is rooted in the sober, steady work of evidence-based risk assessment, targeted interventions, and a long-term focus on building societal resilience. Instead of the all-or-nothing logic of madness, it embraces proportionality, recognizing that not all AI systems pose the same level of risk, and that regulatory attention should be focused where the potential for harm is most significant.

The Mental Shift: Evidence Over Emotion

Moving beyond techno-solutionism means insisting on rigorous evidence before, during, and after deploying AI. It means treating technological solutions with the same skepticism applied to other high-stake poli-

cy interventions. If we are serious about abandoning mythological thinking, we need better tools when asking questions such as “Does this work?” “For whom?” “And at what cost?” Such an evidence-based approach helps us distinguish genuine breakthroughs from overhyped solutions.

The medical field provides a useful model. We do not deploy a new drug to millions based on a marketing pitch. New medical technologies must demonstrate safety and efficacy through controlled trials before receiving regulatory approval.³⁰ In the U.S., medical AI systems that receive approval from the Food and Drug Administration (FDA) must demonstrate effectiveness through controlled trials, leading to genuinely useful applications such as diabetic retinopathy screening and certain forms of medical imaging analysis.³¹ However, many healthcare AI applications are deployed without rigorous evaluation, particularly in areas like clinical decision support and administrative automation.³² Patients deserve better. For example, a Google AI system for retinal disease detection performed accurately during development but exhibited poor performance after deployment in Thai hospitals due to differences in image quality between the training and practice data.³³ Similar standards must apply to AI systems deployed in other critical domains.

IBM Watson Health’s ultimate dissolution after billions in investment demonstrates how promises of salvation can consume enormous resources while delivering minimal clinical benefit.³⁴ The project failed not due to a lack of technology, but rather from unrealistic

promises that overlooked the complexities of medical decision-making. In contrast, DeepMind's AlphaFold was a genuine breakthrough that solved a 50-year-old challenge in protein folding, opening new frontiers in medicine.³⁵ It was technically sophisticated and practically valuable. This is a world away from most viral chatbots, whose technical sophistication is impressive but whose social value is often unclear. In fact, most AI chatbots solve problems that are already adequately addressed by existing technologies, such as search engines, customer service systems, and educational resources; however, they also create problems related to misinformation, privacy, and dependency.

Sometimes, the result can be mixed. When the COVID-19 pandemic upended livelihoods in Togo, the government launched the Novissi program. It set a specific goal to use satellite data and machine learning to improve the targeting of emergency cash assistance.³⁶ The system achieved a meaningful 17-point increase in accuracy, while also being honest about its limitations, including a 30% error rate in excluding truly poor individuals who did not match the algorithmic profile.³⁷ This transparency enabled productive discussions about how AI targeting could supplement, rather than replace, traditional poverty assessment methods.

As a rule of thumb, the goal of smart policy should not be to eliminate enthusiasm for technological innovation, but to channel it toward applications that demonstrably improve human welfare. It means fostering an environment where the next AlphaFold is celebrated as a greater achievement than the next viral chatbot.

The Legal Framework: Adaptive and Risk-Proportional Governance

If moral panic is a disease of governance, then a sense of proportion and timing is the cure. For any government faced with a new technology, the hardest question is “when to act?” Acting too soon on panic risks stifling innovation; acting too late risks allowing harm to take root.

The legislative rush to address deepfakes is a perfect lesson. Initial panic-driven laws in the U.S. focused on the hypothetical threat of video fakes in elections. When the real threat emerged years later as targeted audio manipulations, those early regulations were a poor fit.

The solution is not inaction, but “adaptive vigilance”: designing laws that can evolve as our understanding of a technology grows. This requires iterative processes, for which we have decades of blueprints. We only need to be wise enough to borrow them. The introduction of automobiles, aviation,³⁸ pharmaceuticals,³⁹ and telecommunications⁴⁰ all involved periods of uncertainty and regulatory experimentation before the right balance was found. Each transition offers valuable lessons about striking a balance between innovation and public safety, mitigating unintended consequences, and adapting governance to technological advancements.

For example, the pharmaceutical industry’s phased approach to drug development is a proven model, born from past tragedies and designed to prevent future ones. Phase I trials test basic safety, Phase II evaluates efficacy, and Phase III conducts large-scale comparisons, all before a drug is approved for public use.⁴¹ In some in-

stances, AI deployment could easily follow similar stages: limited testing in controlled environments, evaluation of harms and benefits, and systematic comparison against existing methods before any wide release.

Automotive safety testing offers another relevant framework. Regulators do not just hope that cars are safe but require manufacturers to prove it through standardized tests. In the United States, the National Highway Traffic Safety Administration's Federal Motor Vehicle Safety Standards⁴² require manufacturers to demonstrate compliance through standardized tests before vehicles can be sold. The standards are not static but evolve over time, based on real-world accident data and new technological capabilities, providing a model for creating adaptive safety requirements for AI.

Canada's Algorithmic Impact Assessment⁴³ requirement offers another model for regulatory approaches that resist hype.⁴⁴ It forces government agencies to evaluate a system's potential effects on accuracy, autonomy, procedural fairness, and human dignity before they launch, rather than simply measuring efficiency gains or cost savings.⁴⁵

Regulatory sandboxes offer an alternative avenue for resisting salvation logic by enabling controlled experimentation under democratic oversight. The UK Financial Conduct Authority's sandbox allows fintech companies to test new products with real customers while maintaining consumer protections and, crucially, generating systematic evidence about their benefits and risks.⁴⁶ This model has been formally adopted by the EU AI Act,⁴⁷ whose Article 57 mandates the creation of

AI regulatory sandboxes in member states. These are defined as controlled environments where innovative AI systems can be developed, tested, and validated—sometimes in real-world conditions—for a limited time under regulatory supervision before being placed on the market. Educational AI systems could be piloted in selected schools with a comprehensive assessment of effects on student learning across different populations. Healthcare AI could be tested in controlled clinical settings with rigorous evaluation of diagnostic accuracy and patient outcomes. The key is to ensure the sandboxes serve democratic learning, not just corporate development, with public reporting of findings and a commitment to adjusting the rules based on evidence, not overoptimistic promises by vested interests.

What all these models share is a commitment to iteration and evidence. They feature reviews, pilot studies, and a willingness to adapt—a far cry from the reactive, all-or-nothing prohibitions born of panic.

A focus on impact also requires a critical shift in what we choose to measure. Much of the current regulatory discussion, particularly in high-level international forums, has been captivated by the idea of using computational power, or “compute,” as a primary proxy for risk. The logic is simple: the more computing power a model requires, the more capable—and therefore potentially more dangerous—it must be. This is the approach taken by the Responsible AI Safety and Education Act (RAISE) in New York,⁴⁸ Senate Bill 53 in California,⁴⁹ and the EU’s AI Act.

However, this is another seductive but flawed simplification. Risk is a product of context, not just capability. A small, computationally inexpensive model used to screen asylum applications or recommend medical treatments can cause immense harm, while a massive, high-compute model used to generate poetry is relatively benign. Focusing on compute as the key regulatory threshold is like judging a vehicle's danger solely on its horsepower, ignoring whether it's being driven at 10 mph in a school zone or 90 mph on a wet road by an inexperienced teenager.⁵⁰ Effective regulation understands that the context of deployment—the who, what, and where of its use—is a far more reliable indicator of risk than any technical specification. This approach is also more durable, as it is not rendered obsolete by inevitable advances in algorithmic efficiency that allow for more capable models to be run with less and less computational power.

The Policy Toolkit

Finally, the principle of proportionality is not just about *reacting* correctly; it is about building a society that is proactively resilient to the storms of hype and panic. In this context, resilience is the direct antidote to the “Madness” myth. It is a society's structural capacity to absorb the shock of technological change without breaking—that is, without succumbing to “governance whiplash” or “policy FOMO”. A brittle system reacts with knee-jerk bans or reckless deregulation; a resilient system responds with a steady, evidence-based process. This is achieved by intentionally creating “buffers,” or institutional shock absorbers, at every level. We must

build resilience into our governing bodies, markets, international relations, and public sphere.

1. Institutional Resilience

Resilience requires independent, non-partisan bodies to provide clear-eyed analysis.

The United States had a model for this in its Congressional Office of Technology Assessment (OTA). From 1972 to 1995, it provided balanced analysis that resisted both industry hype and activist alarmism.⁵¹ The OTA covered a variety of emerging technologies, from genetic engineering to computer security, helping legislators understand complex issues without relying solely on the input of interested parties.⁵² While the office was ultimately defunded due to political pressures, its work demonstrated how independent technical expertise can support democratic decision-making about complex technologies.

The U.S. Congressional Budget Office (CBO) offers another model for independent analysis that provides factual foundations for policy debates.⁵³ CBO's economic projections and policy analysis, while not always accurate, create common ground for legislative discussions by establishing shared factual baselines. A modern equivalent, staffed with independent experts, could provide a trusted diagnosis of AI risks, free from the distortions of hype.

2. Market Resilience

Resilience implies aligning market incentives with public values. This means designing systems that reward companies for responsible behavior.

The FDA's "breakthrough device" designation provides a model for incentivizing transparency and responsible innovation.⁵⁴ Companies that demonstrate their devices address unmet medical needs and provide evidence of potential benefits receive expedited review and ongoing communication with FDA staff.

Environmental disclosure requirements offer another precedent for transparency incentives. The United States Environmental Protection Agency's Toxic Release Inventory requires companies to publicly report releases of toxic chemicals, creating market pressures for pollution reduction without mandating specific technologies.⁵⁵

Financial services also demonstrate how transparency requirements can drive industry-wide improvements. The Basel III capital requirements require banks to publicly disclose their risk management practices, creating market incentives for better governance while enabling regulatory oversight.⁵⁶

Similarly, for AI, transparency should be a competitive advantage, not a regulatory burden. Regulators can offer faster approvals for firms that open their models to rigorous testing and monitoring, while subjecting opaque players to heightened scrutiny. AI disclosure requirements could create competitive advantages for companies with better algorithmic governance practices while exposing those with inadequate practices.

3. Global Resilience

We must move beyond the "arms race" mentality (a key part of the Madness myth) and embrace international

cooperation. Challenges such as AI safety, ethics, and economic disruption require coordinated responses. The Montreal Protocol is an inspiring example of successful international technology cooperation.⁵⁷ When scientists identified the threat of ozone depletion from chlorofluorocarbons (CFCs), countries didn't compete to develop the most effective anti-ozone-destroying chemicals. Instead, they worked together to phase out CFCs and develop alternatives. The result was a combination of environmental protection and technological innovation that benefited all participants.

International cooperation is also essential in AI governance, especially in safety research, technical standards, and the management of economic disruption. The Global Partnership on AI represents early efforts to facilitate collaboration, though more structured approaches may be needed as AI technologies become increasingly pervasive.⁵⁸ Learning from successful international frameworks, such as aviation safety coordination and pharmaceutical regulatory harmonization, could inform more effective global approaches to AI governance and promote shared standards.

Yet even as the benefits of international cooperation become apparent, a counter-narrative has emerged that threatens to undermine collaborative approaches: the rhetoric of digital sovereignty. The EU has been particularly vocal in promoting this concept, framing technological independence as essential for political autonomy and economic security.⁵⁹ The move is a response to a legitimate fear of becoming a digital colony, dependent

on U.S. and Chinese platforms for critical infrastructure and data processing.⁶⁰

The digital sovereignty agenda reflects legitimate concerns about technological dependence, but the rhetoric often slips into zero-sum thinking that mirrors the arms-race mentality it claims to reject.⁶¹ When Ursula von der Leyen, President of the European Commission, pledged in 2014 that Europe must achieve “technological sovereignty” in critical areas such as AI,⁶² she framed international cooperation as a luxury that Europe can ill afford rather than a necessity for addressing global challenges.

Such rhetoric creates its own competitive dynamics. If Europe seeks technological autonomy, other regions inevitably respond with their own sovereignty initiatives, fragmenting global technology ecosystems in ways that benefit no one. The result risks delivering not genuine independence but an expensive duplication of effort and a reduced collective capacity to address global challenges.

4. Public Resilience

Public resilience requires a commitment to critical thinking education, empowering citizens to evaluate information independently, understand how technological systems function, and participate in democratic processes that shape the governance of technology.⁶³

The most effective long-term strategy implies a sustained, society-wide investment in robust digital and AI literacy, enabling society to navigate a complex world with wisdom and agency.⁶⁴ It means investing in insti-

tutional strength, ensuring schools, healthcare systems, and community groups are well-funded and accountable, so they can adapt to technological developments while maintaining their core purposes.⁶⁵

It also depends on fostering diverse media ecosystems by supporting public media and local journalism. Such diversity contributes to resilience by ensuring that communities have access to multiple information sources rather than relying on a single platform or algorithm for news and communication.⁶⁶ More importantly, it provides an alternative to algorithmically optimized outrage and empowers citizens to distinguish between legitimate excitement and manipulative hype, building a collective immunity to madness.

Finally, resilience is sustained by developing the social capital and democratic participation infrastructure that allows communities to organize effectively around shared interests rather than remaining isolated and vulnerable to technological manipulation. It, in turn, enables communities to participate in finding adequate solutions to their problems, rather than waiting for a technological savior or fearing a digital demon.⁶⁷

Table 1—The Guardrails Framework

1. CORE PRINCIPLE Categorize Systems by Demonstrated Impact, Not Hypothetical Fantasy RATIONALE & CORRECTIVE FUNCTION Illustration: EU AI Act model that differentiates the obligations of, for example, high-risk medical devices vs. low-risk entertainment recommendations.⁶⁸ Advantage: Forces policymakers to address verifiable, real-world harms, not sci-fi scenarios. Focuses oversight where it is most needed.
2. CORE PRINCIPLE Require Progressively Rigorous Testing Based on Potential Harm RATIONALE & CORRECTIVE FUNCTION Illustration: The FDA's approval process for medical devices, which requires progressively rigorous testing based on the severity of potential patient harm.⁶⁹ Advantage: Acts as a crucial brake on reckless hype cycles, providing a rational alternative to moral panics.
3. CORE PRINCIPLE Mandate Systematic Hazard Identification & Human Accountability RATIONALE & CORRECTIVE FUNCTION Illustration: The FAA's Safety Management System in the aviation sector, which requires airlines to systematically identify hazards, assess risks, and implement controls proportional to an accident's potential severity and likelihood.⁷⁰ Advantage: Restores clear lines of responsibility, making it impossible to blame the algorithm for failures.

4. CORE PRINCIPLE

Incentivize Safety by Making Risk a Material Cost

RATIONALE & CORRECTIVE FUNCTION

Illustration: Financial regulation models, such as the Basel framework, require banks to hold capital proportional to their assessed risks.

Advantage: Replaces the need for blind faith in corporate benevolence with verifiable, economic incentives, forcing companies to internalize the costs of the risks their systems create.

Principle#3: Durability as Antidote to Heaven

The final principle for mature governance is durability. This approach conceives of AI regulation not as a series of reactive, technology-specific laws, but as a set of foundational principles. It is the direct antidote to the opposing temptations of the Heaven myth, which urges a dangerous deference to tech prophets, and the Sin myth, which demands reactionary prohibitions.

The Mental Shift: From Reactive Laws to Durable Principles

Instead of writing piecemeal legislation for each new technological marvel or abdicating governance entirely, this framework establishes a stable and future-proof set of high-level principles to guide all algorithmic systems.

Durability requires a shift in our long-term thinking. While the tool metaphor is useful for accountability, the infrastructure analogy is better for understanding AI's systemic influence. Like electricity or transportation networks, AI is becoming a foundational layer of society that requires robust governance frameworks,

clear technical standards, equitable access provisions, and comprehensive safeguards against abuse.⁷¹ This framing forces us to consider public-interest principles such as universal access, reliability, and interoperability.

Additionally, the infrastructure approach alters our perspective on the concentration of the AI market. Rather than treating AI capabilities as commercial products subject only to market forces, infrastructure thinking acknowledges that access to AI tools is increasingly influencing opportunities for education, employment, and civic participation. Just as telecommunications and transportation infrastructure are subject to common carrier regulations and antitrust oversight, AI infrastructure might require similar governance mechanisms to prevent monopolization and ensure competitive markets.

Resilience and reliability become increasingly crucial as AI systems become more deeply integrated into critical social systems. Infrastructure frameworks emphasize the need for redundancy, fail-safe mechanisms, and disaster recovery planning. Such concepts translate naturally to AI governance through requirements for algorithmic auditing, backup systems, and transparency mechanisms that enable oversight and accountability.

Finally, a durable framework recognizes that technical fixes alone are insufficient. Many issues framed as “AI problems” are actually systemic, human problems amplified by new tools. A holistic approach requires embracing “complexity thinking”: seeing social challenges not as bugs to be fixed but as outcomes of tan-

gled systems, histories, and power dynamics.⁷² Poverty, for instance, is not a data distribution issue; it is a convergence of policy, labor, education, and structural injustice. Climate change is not just a technological problem; it requires coordinated changes in energy systems, transportation, and international cooperation. Technology can play a supporting role, but it cannot substitute for political will or public buy-in.

Complexity thinking looks beyond the algorithm to the broader social, economic, and political systems. Achieving this requires cross-sector collaboration, bringing together technologists, social scientists, educators, and community organizations to develop comprehensive strategies.⁷³

- When we blame automation for unemployment, we must also look at economic policy, job training, and stronger safety nets.⁷⁴
- When we blame platforms for loneliness, we must also support community development, public spaces, civic organizations, and programs that foster the face-to-face connections that no technology can replace.
- When we blame policy failures on inscrutable algorithms, we must invest in the cross-disciplinary research needed to build a comprehensive picture of their impacts, moving beyond the isolated perspectives of computer science or social science.

The Legal Framework: Principles-Based Regulation

The model for durability draws inspiration from the success of Principles-Based Regulation (PBR) in other complex, fast-moving domains. In fields such as finance

and environmental protection, experience reveals that attempting to write thousands of prescriptive rules is not a good idea, as they are likely to become obsolete quickly.⁷⁵ PBR, by contrast, sets broad, outcome-focused standards. For example, the UK FCA imposes a duty on firms to “act to deliver good outcomes for retail customers.”⁷⁶ This approach endures in the face of innovation, forcing firms to engage with the spirit of the law and reducing the potential for creative compliance.

The Uniform Commercial Code (UCC) in the United States stands as a testament to this thinking. For over seventy years, the UCC has provided a stable and uniform legal foundation for commercial transactions, successfully adapting to the shift from an industrial economy to the era of e-commerce.⁷⁷ It achieves this durability not through detailed technological specifications but through clear, high-level principles governing contracts, sales, and secured transactions, successfully adapting from an industrial economy to the age of cryptocurrency.

An AI PBR would similarly establish a set of foundational, technology-neutral tenets:

- Human agency and oversight: A default requirement for meaningful human control over consequential decisions.
- Accountability and responsibility: A clear allocation of liability to the human actors who design, deploy, and profit from AI systems.

- Transparency and explainability: A right for individuals to understand the basis of algorithmic decisions that affect them.
- Fairness and non-discrimination: A prohibition on algorithmic systems that produce or perpetuate unlawful discrimination.
- Safety and reliability: A duty to ensure that AI systems operate safely and reliably in their intended contexts.

The Policy Toolkit: Reclaiming Democratic Control

After detailing the pervasive capture of our democratic institutions, is it not naive to propose democratic realism as the solution? This is not a contradiction; it is the core of the challenge. The fact that our institutions are under strain is not a reason to abandon them for technocratic prophets. On the contrary, it is the reason we must work to reclaim and strengthen them.

1. Tools for Democratic Participation

We already have powerful models for structured public engagement. As noted in Chapter 4, Ireland's Citizens' Assembly demonstrates that a representative group of everyday people, given time and access to independent expertise, can produce nuanced policy recommendations on deeply divisive issues such as abortion and climate change.⁷⁸ This same approach can be applied to AI governance, replacing the polarized debate of salvation versus damnation with evidence-based discussion. Imagine citizen assemblies systematically tackling the questions that the salvation narrative ignores. Should public schools use AI tutoring systems, and under what conditions? How should healthcare AI be designed to

preserve patient privacy? What role should algorithms play in the justice system, and what human oversight is essential? These are precisely the kinds of messy, value-laden debates that the salvation narrative seeks to override but that democracy is designed to navigate.

The Danish Board of Technology's consensus conferences have equally provided a model for public engagement on technological issues.⁷⁹ These conferences bring together diverse citizens to examine emerging technologies, hear from independent experts, and develop recommendations that have successfully influenced democratic policy.⁸⁰

We even have examples of how digital tools can support this process, such as the vTaiwan platform mentioned in Chapter 3, which has enabled large-scale public consultation on complex issues such as ride-sharing and cryptocurrency regulation, using technology to enhance, not replace, democratic deliberation.⁸¹

2. Tools for Transparency and Agency

For democratic participation to be meaningful, it must be informed. It requires tools that empower citizens and embed human agency into the design of technology itself.

First, transparency must be designed for public understanding, not just for bureaucratic box-checking. Public registries of automated decision-making systems, such as those in Amsterdam⁸² and Helsinki, are a crucial step.⁸³ They require government agencies to disclose how and when they use algorithmic systems in clear, accessible language, not technical jargon. The

Amsterdam registry details each system's purpose, decision-making logic, data sources, accuracy metrics, and human oversight procedures, allowing citizens to understand and evaluate the automated decisions that affect their lives and communities. The Helsinki registry goes a step further by requiring public consultation before deploying AI systems that could significantly affect citizen rights.⁸⁴ This consultation process creates opportunities for community input from the outset on whether algorithmic applications serve public purposes, while also enabling ongoing oversight of system performance and social impacts.

Second, transparency frameworks must also create a legal right to an explanation for citizens affected by automated decisions. France's algorithmic transparency law, for instance, allows citizens affected by automated government decisions about benefits, taxes, or other government services to demand an accessible explanation of how those decisions were reached, establishing a framework for accountability that goes beyond mere technical disclosure.⁸⁵ Corporate transparency requirements could follow similar principles by requiring companies to disclose the impacts of AI systems on workers, consumers, and communities in a meaningful manner. The EU's AI Act includes transparency requirements for certain AI applications, but these could be expanded to cover broader social and economic impacts that affect public welfare.⁸⁶

Third, beyond transparency, we need community-controlled technology. Barcelona's Decidim offers an example of participatory democracy software designed

by and for civic organizations, rather than by a tech company optimizing for engagement.⁸⁷ The platform enables neighborhood assemblies, municipal councils, and civic organizations to organize democratic deliberation. Its governance is democratic, and its success in supporting participatory budgeting and neighborhood planning demonstrates how technology can serve public purposes. Cooperative ownership models, such as platform cooperatives, offer a similar path, enabling communities to collectively own and govern the systems that impact their lives. Platform cooperatives such as Stocksy United and Resonate demonstrate how worker and user ownership can align technological development with community interests, moving beyond a singular focus on profit maximization.⁸⁸ Public banking provides a related model for funding such initiatives, allowing community-controlled financial institutions to support local, democratic technology development, free from the commercial pressures that fuel the salvation myth.⁸⁹

Fourth, we must build widespread technological literacy. This means rejecting the narrative of helplessness. Reclaiming agency begins with education that frames AI as a human creation that can be shaped through democratic engagement. Finland, for example, launched a program in 2020 to teach at least 1% of its citizens—and, by extension, civil servants—the basics of AI.⁹⁰ The campaign quickly exceeded its targets and offered a model for broader digital education.⁹¹ The idea is simple: if lawmakers, judges, city council members, and voters understand what AI can and cannot

do, they are better equipped to ask tough questions.⁹² Similarly, when workers can see the logic behind algorithmic management, they are better equipped to organize for fair treatment and oversight.⁹³ This literacy must extend to computer science education, which should include the analysis of algorithmic bias and ethical design, not just technical implementation.⁹⁴

Finally, we must design systems that preserve human agency. This requires creating interfaces that allow users to understand and modify algorithmic decisions, providing options for human oversight, and maintaining human-centered workflows that utilize technology to augment, rather than replace, human judgment.⁹⁵

3. Tools for Power-Balancing

More generally, a durable democracy must build structural resistance to myth-based capture. This means rebalancing power.

Media: Restructure the Pulpit

Technology journalism plays a crucial role in shaping public and policy understanding of AI, through editorial choices about coverage, sourcing, and framing that influence whether industry narratives are accepted uncritically or subjected to rigorous evaluation. It often amplifies industry perspectives rather than providing independent analysis. As the book *AI Snake Oil* notes, “If researchers and companies kindle the sparks of hype, the media fans the flames.”⁹⁶

This amplification isn’t limited to industry hype; it extends to bizarre political theater. In October 2025, when Albania’s Prime Minister announced his AI “min-

ister” was “pregnant” with 83 digital assistants,⁹⁷ he was framing a simple software rollout as a magical birth. Instead of ignoring this bizarre take, the media picked it up as a journalist at a European Commission briefing asked for a reaction to the happy news,” and the Commission’s spokesperson, rather than correcting the absurd anthropomorphism, played along, offering official “congratulations.”⁹⁸ This is a perfect example of the media and policy institutions choosing to amplify a narrative rather than ignoring it or providing the critical analysis needed to hold power to account. Fostering a media ecosystem capable of more independent, critical analysis is therefore essential. Equally important in today’s diverse media environment is investing in public media literacy, enabling citizens to critically assess the technological narratives presented to them. Fostering a resilient media ecosystem, such as by supporting public media or initiatives like the BBC’s Verify service,⁹⁹ is crucial for countering misinformation and myths.

Research: Open the Black Box

AI research and development are often conducted in the laboratories of a few corporate giants, shielded from democratic oversight despite being fueled by public investment in the form of tax incentives, research grants, and infrastructure support. If democracy is to have any say in our technological future, it must have a seat at the table where that future is being designed.

Further, public investment must be used to create alternatives to corporate-controlled technology. When governments fund AI research, they can attach strings that

require public benefit, open development, and democratic accountability, creating a technology ecosystem that serves more than commercial advantage.¹⁰⁰

Finally, academic institutions require structural reforms that reduce dependence on industry funding. They must be bastions of independent thought, not extensions of corporate R&D. Even though it is currently under pressure due to massive budget cuts from the Trump administration,¹⁰¹ the National Science Foundation's historic support for fundamental research provides a model for public investment that generates knowledge serving broad social interests rather than narrow commercial applications.¹⁰² The goal is not to eliminate industry collaboration but to ensure that academic research maintains the independence necessary for genuine evaluation of technological benefits and risks.

International cooperation, guided by frameworks such as the Montreal Declaration for Responsible AI,¹⁰³ can further align research priorities with shared democratic values, resisting the pull of a competitive race to the bottom in favor of a collaborative pursuit of the public good. Democratic countries could collaborate on research priorities that advance shared values of human rights, democratic governance, and social justice, instead of leaving AI governance to competition between national technology champions.

Market: Address Concentration

The extreme concentration of wealth and market power in the hands of a few tech companies warrants scrutiny. Robust antitrust enforcement is crucial, as it addresses the market concentration that enables these companies

to shape public discourse and eliminate competitive alternatives. When a few companies control the core infrastructure of AI development, their leaders naturally become like prophets, regardless of the merit of their visions. Addressing the potential negative impact of such concentration is a prerequisite for breaking the spell of the Heaven myth.

A Coherent System for Governance

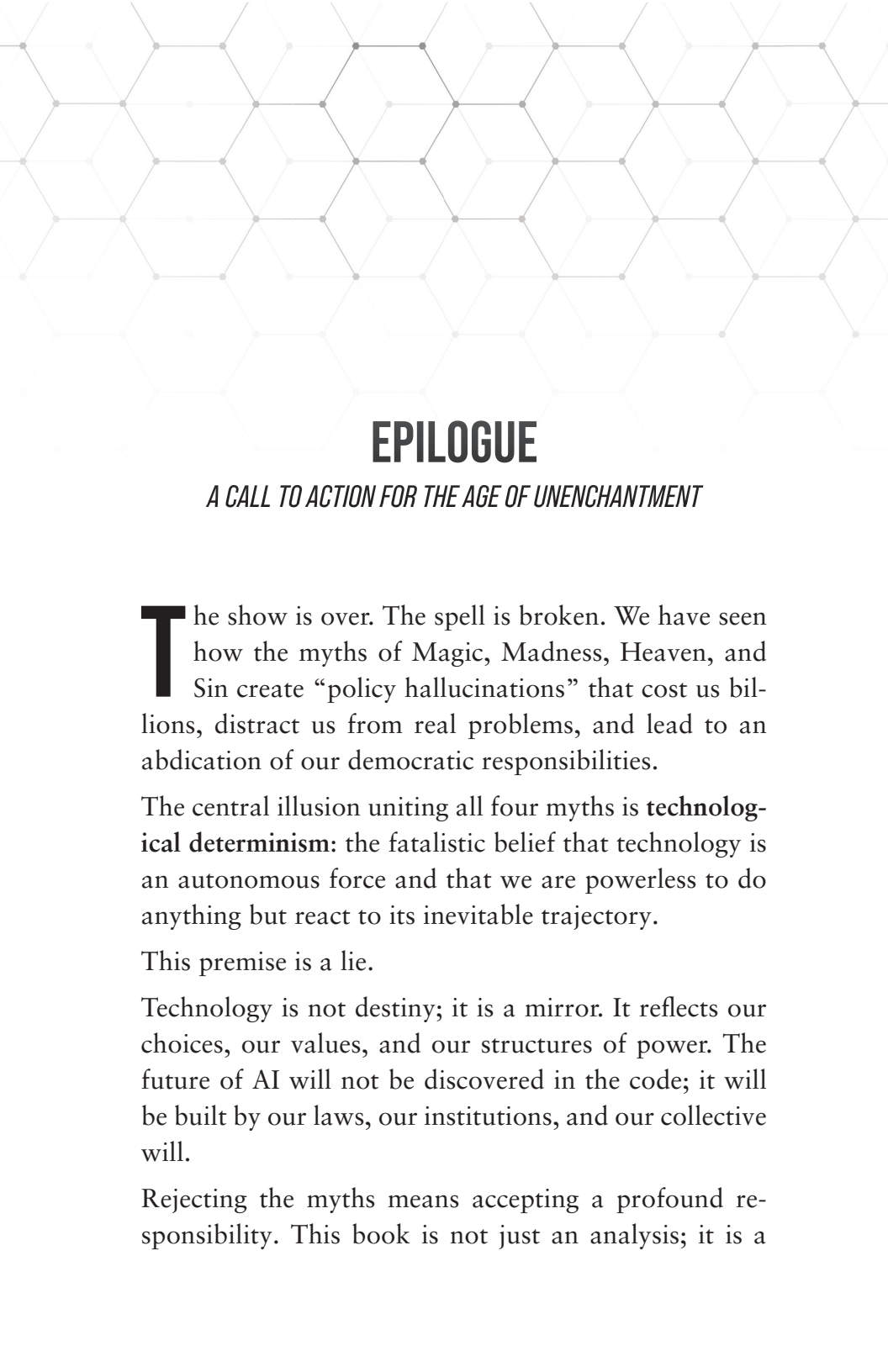
The three frameworks of Accountability, Proportionality, and Durability are complementary layers of a coherent system for mature governance. The AI as Product framework provides the necessary sword, establishing clear liability for the platform's own actions and designs. The Public Health model creates the expert agencies needed to assess risk and inform both regulators and the public. Finally, the Principle-Based Regulation approach provides the overarching values that guide the system. It is a system of checks and balances for the algorithmic age, one that can simultaneously protect expression and hold powerful actors accountable for the harms their systems directly cause.

Ultimately, while the catalyst for this analysis has been the rise of AI, the principles of accountability, proportionality, and durability outlined are not unique to AI. They are the essential components of wise governance. The next transformative force may emerge from biotechnology, quantum computing, or a field yet to be imagined. If we learn the lessons of the AI era, we can build a resilient democratic framework ready for that

future: one that governs based on evidence and shared values, not on myth and fear.

Table 2—Flawed Mythologies vs.
the Mature Governance Frameworks

MAGIC (Mystification) RESULTING POLICY FAILURE Policy Distraction & Techno-Solutionism MATURE GOVERNANCE PRINCIPLE Accountability CORRESPONDING FRAMEWORK AI as a Product: Demystifies AI by applying product liability law, focusing on concrete harms and manufacturer responsibility.	MADNESS (Hype/Panic) RESULTING POLICY FAILURE Emergency Governance & Policy Whiplash MATURE GOVERNANCE PRINCIPLE Proportionality CORRESPONDING FRAMEWORK AI Governance as Public Health: Replaces emotional extremes with evidence-based risk assessment and targeted interventions.
HEAVEN (Salvation) RESULTING POLICY FAILURE Prophetic Deference & Democratic Abdication MATURE GOVERNANCE PRINCIPLE Durability CORRESPONDING FRAMEWORK AI Regulation as PBR: Establishes stable, high-level principles to prevent capture by utopian promises.	SIN (Scapegoating) RESULTING POLICY FAILURE Blame Deflection & Flawed Regulation MATURE GOVERNANCE PRINCIPLE Shared Responsibility CORRESPONDING FRAMEWORK Integrated System: Holds human actors accountable (via liability) instead of blaming bots, while protecting platforms as forums.



EPILOGUE

A CALL TO ACTION FOR THE AGE OF UNENCHANTMENT

The show is over. The spell is broken. We have seen how the myths of Magic, Madness, Heaven, and Sin create “policy hallucinations” that cost us billions, distract us from real problems, and lead to an abdication of our democratic responsibilities.

The central illusion uniting all four myths is **technological determinism**: the fatalistic belief that technology is an autonomous force and that we are powerless to do anything but react to its inevitable trajectory.

This premise is a lie.

Technology is not destiny; it is a mirror. It reflects our choices, our values, and our structures of power. The future of AI will not be discovered in the code; it will be built by our laws, our institutions, and our collective will.

Rejecting the myths means accepting a profound responsibility. This book is not just an analysis; it is a

call to action. The principles of Accountability, Proportionality, and Durability are not abstract theories; they are a blueprint for the patient, necessary, and essential work of democratic governance.

This work belongs to all of us.

To the Policymaker

Your desk is littered with proposals that either ban AI in a panic or deregulate it in a fever. You are being pushed and pulled by the myths of Madness and Sin.

Resist them. Your task is not to regulate “AI” as a magical, monolithic entity. Your task is to govern specific, high-risk activities enabled by these tools. Embrace **technology-neutrality**. Instead of writing a law for “ChatGPT,” strengthen the laws against fraud, discrimination, and non-consensual image abuse, no matter what tool is used.

Stop legislating for Merlin. Adopt the “AI as Product” framework. Hold companies liable for their defective designs, just as you would an automaker. Adopt the “Public Health” model. Focus regulatory energy proportionally on high-risk applications, such as algorithms used in medicine and law, rather than on low-risk tools, like spam detection.

Do not abdicate your duty to the prophets. The future is not their vision to bestow; it is your responsibility to build.

To the Journalist

You are the first line of defense against the myths. Yet, the “Policy Machine” depends on media to amplify hype and panic.

Reject the script. When a tech CEO makes a grand promise (the **Heaven** myth) or a dire warning (the **Sin** myth), do not report it as a weather forecast. Treat it as what it is: a sophisticated press release.

Ask the hard questions. Move past the “magic” of the demo and ask about the mechanics. Ask about the training data, the labor conditions of the annotators, and the documented error rates. Ask for evidence, not just evangelism. When you hear the words “inevitable,” “sentient,” or “superintelligence,” your skepticism should be at its peak.

Your job is not to transcribe the prophecies; it is to investigate the prophets.

To the Citizen, the Educator, and the Leader

You are told that this technology is too complex, too fast, and too powerful for you to understand. This is the **Magic** myth at its most disempowering.

Reclaim your agency. You do not need to be a data scientist to have a legitimate voice in this debate. You are an expert in what it means to be human, and that is the only expertise that matters in a democracy.

Demand a seat at the table. Organize. Join citizens’ assemblies. Ask your local school board, your city council, and your employer how they are using automated systems. Demand transparency, not in code, but in plain language. Ask: “What is this tool for?” “How does it make decisions?” “Who is responsible when it is wrong?”

The challenge of AI is not a technical problem for engineers to solve. It is a political and social challenge

that calls for the best of our democratic traditions. This epilogue is a call to end the age of enchantment and embrace the necessary work of building the future we want, not the one we've been sold.

The burden is heavy, but the promise is a future that serves human flourishing, not the other way around. That is the burden. That is the promise. Now, let's get to work.

ENDNOTES

Introduction

- 1 Abraham Joshua Heschel, *Moral Grandeur and Spiritual Audacity: Essays*, ed. Susannah Heschel (New York: Farrar, Straus and Giroux, 1997), <https://archive.org/details/moralgrandeurspi0000hesc>
- 2 George Lakoff, *Don't Think of an Elephant! Know Your Values and Frame the Debate* (White River Junction, VT: Chelsea Green Publishing, 2004), <https://chelseagreen.com/product/dont-think-of-an-elephant/>
- 3 Karen Hao, *Empire of AI* (Allen Lane, 2025), <https://www.penguin.co.uk/books/460331/empire-of-ai-by-hao-karen/9780241678923>
- 4 Artificial Intelligence (AI) Coined at Dartmouth,” Dartmouth Today, Dartmouth College, <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>
- 5 Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton, NJ: Princeton University Press, 2024), <https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil>
- 6 Wikipedia, s.v. “Large language model,” https://en.wikipedia.org/wiki/Large_language_model

Chapter 1

- 1 Andrew P. Morriss, “The Wild West Meets Cyberspace.” Foundation for Economic Education, n.d. <https://fee.org/articles/the-wild-west-meets-cyberspace/>
- 2 Unknown author (username rbanffy), “Crypto’s Wild West Era Is over | Hacker News,” n.d. <https://news.ycombinator.com/item?id=44603096>
- 3 Thierry Breton (@ThierryBreton), “It’s time to put some order in the digital “Wild West”. A new sheriff is in town — and it goes by the name #DSA,” X (formerly Twitter), January 19, 2022, <https://x.com/ThierryBreton/status/1483786510214303744>
- 4 Vinton G. Cerf, “Our 21st Century Information Superhighway,” *Ingenia* (adapted from the print edition of *Ingenia* no. 82, March 2020), <https://www.ingenia.org.uk/articles/our-21st-century-information-superhighway/>
- 5 Brewster Kahle, “Imagining the Internet: Explaining Our Digital Transition | Brewster Kahle’s Blog,” January 13, 2022. <https://brewster.kahle.org/2022/01/13/imagining-the-internet-explaining-our-digital-transition/>
- 6 Clay Calvert, “Regulating Cyberspace: Metaphor, Rhetoric, Reality, and the Framing of Legal Options,” *Hastings Communications and Entertainment Law Journal* 20, no. 3 (1998): 541–65, https://repository.uclawsf.edu/hastings_comm_ent_law_journal/vol20/iss3/3
- 7 Wikipedia, s.v. “Information superhighway,” https://en.wikipedia.org/wiki/Information_superhighway
- 8 Kristen Jakobsen Osenga, “The Internet Is Not a Super Highway: Using Metaphors to Communicate Information and Communications Policy,” UR Scholarship Repository, <https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1047&context=law-faculty-publications>
- 9 Will Davis, “Eisenhower’s Atomic Power for Peace – The Civilian Application Program and Power Demonstration Reactor,” *ANS Nuclear Newswire*, March 20, 2014, <https://www.ans.org/news/article-1486/atomic-power-for-peace-the-civilian-application-program-and-power-demonstration-reactor/>
- 10 Hanani Hlomani et al., “Dissent and Resistance to Silicon Valley AI Narratives,” *TechPolicy.Press*, October 24, 2023, <https://www.techpolicy.press/dissent-and-resistance-to-silicon-valley-ai-narratives/>
- 11 Philip Rossetti, “What Happened to Electricity Becoming ‘Too Cheap to Meter?’” *The Daily Dish* (blog, American Action

- Forum), August 30, 2017, <https://www.americanactionforum.org/daily-dish/happened-electricity-becoming-cheap-meter/>
- 12 Barbara Gabriella Renzi, Matthew David Cotton, Ralf Barkemeyer, and Giulio Napolitano, "Rebirth, Devastation and Sickness: Analyzing the Role of Metaphor in Media Discourses of Nuclear Power," *Environmental Communication* 11, no. 5 (March 22, 2016): 624–40, <https://doi.org/10.1080/17524032.2016.1157506>
 - 13 Benoît Pelopidas, Pelopidas, Benoît. "The Oracles of Proliferation" *The Nonproliferation Review* 18, no. 1 (February 19, 2011): 297–314. <https://doi.org/10.1080/10736700.2011.549185>
 - 14 Michael A. St. Jacques, "Their Swords, Our Plowshares: 'Peaceful' Nuclear Weapons, Propaganda, and Cold War Memory Expressed in Film: 1959-1989." *JMU Scholarly Commons*, n.d. <https://commons.lib.jmu.edu/master201019/102/>
 - 15 Seth C. Lewis, David M. Markowitz, and Jon Benedik Bunquin. "Journalists, Emotions, and the Introduction of Generative AI Chatbots: A Large-Scale Analysis of Tweets before and after the Launch of ChatGPT." *arXiv.org*, September 13, 2024. <https://arxiv.org/abs/2409.08761>
 - 16 Matthew Burtell and Helen Toner, "The Surprising Power of Next Word Prediction: Large Language Models Explained, Part 1," *Center for Security and Emerging Technology*, March 8, 2024, <https://cset.georgetown.edu/article/the-surprising-power-of-next-word-prediction-large-language-models-explained-part-1/>
 - 17 Melanie Mitchell, "The Metaphors of Artificial Intelligence," *Science* 383, no. 6687 (March 7, 2024): 1031, <https://doi.org/10.1126/science.adt6140>
 - 18 Luciano Floridi and Anna C. Nobre, "Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences," *Minds and Machines* 34, no. 1 (April 25, 2024): 1–9, <https://doi.org/10.1007/s11023-024-09670-4>
 - 19 Leon Furze, "AI Metaphors We Live By: The Language of Artificial Intelligence," *Leon Furze*, July 19, 2024, <https://leonfurze.com/2024/07/19/ai-metaphors-we-live-by-the-language-of-artificial-intelligence/>
 - 20 Matthijs Maas, "AI is like....: A Literature Review of AI Metaphors and Why They Matter for Policy," *Institute for Law & AI, AI Foundations Report 2*, October 2023, <https://law-ai.org/ai-policy-metaphors/>

- 21 Douglas Galbi, Review for EH.NET of “Paul Starr, *The Creation of the Media: Political Origins of Modern Communications*,” EH.NET, https://eh.net/book_reviews/the-creation-of-the-media-political-origins-of-modern-communications/
- 22 Adriana Placani, “Anthropomorphism in AI: Hype and Fallacy,” *AI and Ethics* 4, no. 3 (February 5, 2024): 691–698, <https://doi.org/10.1007/s43681-024-00419-4>
- 23 Brigitte Nerlich, “Hunting for AI Metaphors,” *Making Science Public* (University of Nottingham blog), April 12, 2024, <https://blogs.nottingham.ac.uk/makingsciencepublic/2024/04/12/hunting-for-ai-metaphors/>
- 24 Wikipedia, s.v. “Futurama”, <https://en.wikipedia.org/wiki/Futurama>
- 25 Gary Marcus, “Deep Learning: A Critical Appraisal,” arXiv preprint, January 2, 2018, <https://doi.org/10.48550/arXiv.1801.00631>
- 26 Madhumita Murgia, interview of Ted Chiang, “Sci-fi writer Ted Chiang: ‘The machines we have now are not conscious’,” *Financial Times*, June 3–4, 2023, <https://www.ft.com/content/c1f6d948-3dde-405f-924c-09cc0dcf8c84>
- 27 Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FACCT ’21)*, March 3–10, 2021 (virtual), 610–623, <https://doi.org/10.1145/3442188.3445922>
- 28 Sheila Jasanoff (Ed.), *States of Knowledge: The Co-Production of Science and the Social Order* (London: Routledge, 2004), <https://doi.org/10.4324/9780203413845>
- 29 Stephen Cave, Claire Craig, Kanta Dihal, Sarah Dillon, Jessica Montgomery, Beth Singler, and Lindsay Taylor. 2018. “Portrayals and Perceptions of AI and Why They Matter.” *The Royal Society*, <https://doi.org/10.17863/CAM.34502>
- 30 Leon Furze, “Myths, Magic, and Metaphors: The Language of Generative AI,” Leon Furze, October 18, 2023, <https://leonfurze.com/2023/10/18/myths-magic-and-metaphors-the-language-of-generative-ai/>
- 31 Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016), <https://www.amazon.com/exec/obidos/ASIN/0553418815/acmorg-20>
- 32 Klaus Schwab, *The Fourth Industrial Revolution* (Geneva: World Economic Forum, 2016), https://law.unimelb.edu.au/__data/

- assets/pdf_file/0005/3385454/Schwab-The_Fourth_Industrial_Revolution_Klaus_S.pdf
- 33 Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, February 2018), 256 pp., <https://nyupress.org/9781479837243/>
 - 34 Matthijs Maas, “AI is like...: A Literature Review of AI Metaphors and Why They Matter for Policy,” Institute for Law & AI, AI Foundations Report 2, October 2023, <https://law-ai.org/ai-policy-metaphors/>
 - 35 Amos Tversky and Daniel Kahneman, “Availability: A Heuristic for Judging Frequency and Probability,” *Cognitive Psychology* 5, no. 2 (September 1973): 207–32, [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)
 - 36 George Lakoff and Mark Johnson, *Metaphors We Live By*, updated edition (Chicago: University of Chicago Press, 2003), 256 pp., <https://press.uchicago.edu/ucp/books/book/chicago/M/bo3637992.html>
 - 37 Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven: Yale University Press, 2006), 515 pp., https://www.benkler.org/Benkler_Wealth_Of_Networks.pdf
 - 38 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), 320 pp., <https://raley.english.ucsb.edu/wp-content/Engl800/Pasquale-blackbox.pdf>
 - 39 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), <https://www.publicaffairsbooks.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395700/>
 - 40 Amba Kak and Sarah Myers West, “A Modern Industrial Strategy for AI?: Interrogating the US Approach,” AI Now Institute, <https://ainowinstitute.org/publication/a-modern-industrial-strategy-for-aiinterrogating-the-us-approach>
 - 41 Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford: Oxford University Press, 2016), <https://doi.org/10.1093/acprof:oso/9780190498511.001.0001>

Chapter 2

- 1 Clothilde Goujard, “Italian Privacy Regulator Bans ChatGPT,” *Politico*, March 31, 2023, <https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/>

- 2 K.C. Halm, John D. Seiver and Patrick J. Austin, "Italy's Data Protection Agency Lifts Ban on ChatGPT," Davis Wright Tremaine – Artificial Intelligence Law Advisor (blog), May 15, 2023, <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/05/ai-chatgpt-italy-ban-lifted>
- 3 Natasha Lomas, "ChatGPT Resumes Service in Italy after Adding Privacy Disclosures and Controls," TechCrunch, April 28, 2023, <https://techcrunch.com/2023/04/28/chatgpt-resumes-in-italy/>
- 4 Frank R. Baumgartner and Bryan D. Jones, *Agendas and Instability in American Politics* (Chicago: University of Chicago Press, 1993), <https://archive.org/details/agendasinstabili0000baum>
- 5 Peter A. Hall, "Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain," *Comparative Politics* 25, no. 3 (April 1993): 275–96, <https://doi.org/10.2307/422246>
- 6 Steven Shapin, and Simon Schaffer, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton: Princeton University Press, 1985), https://openlibrary.org/books/OL21623774M/Leviathan_and_the_air-pump
- 7 Julie Thompson Klein, *Interdisciplinarity: History, Theory, and Practice* (Detroit: Wayne State University Press, 1990), <https://archive.org/details/interdisciplinar00kleirich>
- 8 Nur Ahmed, Muntasar Wahed, and Neil C. Thompson, "The Growing Influence of Industry in AI Research," *Science* 379, no. 6635 (March 3 2023): 884–886, <https://doi.org/10.1126/science.ade2420>
- 9 Karen Hao, *Empire of AI* (Allen Lane, 2025), <https://www.penguin.co.uk/books/460331/empire-of-ai-by-hao-karen/9780241678923>
- 10 Maja Schyns, "Why and how is the power of Big Tech increasing in the policy process? The case of generative AI," *Policy and Society* 44, no. 1 (2025): 52–73, <https://academic.oup.com/policyandsociety/article/44/1/52/7636223>
- 11 Diana Crane, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (Chicago: University of Chicago Press, 1972), <https://archive.org/details/invisiblecollege0000cran>
- 12 Andrew Rich, *Think Tanks, Public Policy, and the Politics of Expertise* (Cambridge & New York: Cambridge University Press, 2004), <https://archive.org/details/thinktankspubli00andr>
- 13 Murray J. Edelman, *Constructing the Political Spectacle* (Chicago: University of Chicago Press, 1988), <https://archive.org/details/constructingpoli00edel>

- 14 Natasha Dow Schüll, "Data for Life: Wearable Technology and the Design of Self-Care," *BioSocieties* 11, no. 3 (October 2016): 317–33, <https://doi.org/10.1057/biosoc.2015.47>
- 15 Bobby Allyn, "Stung By Media Coverage, Silicon Valley Starts Its Own Publications," NPR, June 25, 2021, <https://www.npr.org/2021/06/25/1010066447/stung-by-media-coverage-silicon-valley-starts-its-own-publications>
- 16 Wikipedia, s.v. "Churnalism," <https://en.wikipedia.org/wiki/Churnalism>
- 17 Matthew Hindman, *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy* (Princeton: Princeton University Press, 2018), <https://doi.org/10.23943/princeton/9780691159263.001.0001>
- 18 Herbert J. Gans, *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time* (Evanston, IL: Northwestern University Press, 2004), https://openlibrary.org/books/OL3312029M/Deciding_what%27s_news
- 19 Madhumita Murgia, Cristina Criddle, and Melissa Heikkilä, "AI pioneers claim human-level general intelligence is already here," *Financial Times*, November 6, 2025, <https://www.ft.com/content/5f2f411c-3600-483b-bee8-4f06473ecd0>
- 20 Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Ashton Schlaikjer, and Moritz Hardt, "Algorithmic Amplification of Politics on Twitter," *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 1 (2022): e2025334119, <https://doi.org/10.1073/pnas.2025334119>
- 21 Benedict R. O'G. Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, rev. ed. (London & New York: Verso, 2006), https://archive.org/details/imaginedcommunit00ande_0
- 22 James Lewis, "AI and Arms Races," Center for European Policy Analysis, June 3, 2025, <https://cepa.org/article/ai-and-arms-races/>
- 23 John W. Kingdon, *Agendas, Alternatives, and Public Policies*, 2nd ed. (New York: HarperCollins College Publishers, 1995), <https://archive.org/details/agendasalternati00king>
- 24 Alvin Carpio, "The rise of the political entrepreneur and why we need more of them," World Economic Forum, November 23, 2017, <https://www.weforum.org/stories/2017/11/the-rise-of-the-political-entrepreneur-and-why-we-need-more-of-them/>
- 25 Deborah Stone, *Policy Paradox: The Art of Political Decision Making*, 3rd ed. (New York: W.W. Norton & Company, 2012), <https://archive.org/details/policyparadoxart0000ston>

- 26 Paul Pierson, "Increasing Returns, Path Dependence, and the Study of Politics," *American Political Science Review* 94, no. 2 (June 2000): 251–67, <https://www.jstor.org/stable/2586011>
- 27 Theda Skocpol, *Diminished Democracy: From Membership to Management in American Civic Life* (Norman, OK: University of Oklahoma Press, 2003), <https://archive.org/details/diminisheddemocr0000skoc>
- 28 John W. Kingdon, *Agendas, Alternatives, and Public Policies*, 2nd ed. (New York: HarperCollins College Publishers, 1995), <https://archive.org/details/agendasalternati00king>
- 29 Ivan Jureta, "Policy Windows: What They Are And When They Occur," Ivan Jureta blog, December 2024, <https://ivanjureta.com/policy-windows-what-they-are-and-when-they-occur/>
- 30 Michael Lipsky, *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services* (New York: Russell Sage Foundation, 1980), <https://archive.org/details/streetlevelburea0000lips>
- 31 James Q. Wilson, *Bureaucracy: What Government Agencies Do and Why They Do It* (New York: Basic Books, 1989), <https://archive.org/details/bureaucracywhatg00wils>
- 32 Cary Coglianese and David Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era," *Georgetown Law Journal* 105, no. 5 (2017): 1147–1223, https://scholarship.law.upenn.edu/faculty_scholarship/1734
- 33 James Q. Wilson, *Bureaucracy: What Government Agencies Do and Why They Do It* (New York: Basic Books, 1989), <https://archive.org/details/bureaucracywhatg00wils>
- 34 Cornelius M. Kerwin and Scott R. Furlong, *Rulemaking: How Government Agencies Write Law and Make Policy*, 4th ed. (Washington, DC: CQ Press, 2010), https://archive.org/details/rulemakinghowgov0000kerw_4edi
- 35 Ronald M. Dworkin, *Law's Empire* (Cambridge, MA: Belknap Press of Harvard University Press, 1986), <https://archive.org/details/details/lawsempire0000dwor>
- 36 Edward H. Levi, *An Introduction to Legal Reasoning* (Chicago: University of Chicago Press, 1949)
- 37 Jerome N. Frank, *Courts on Trial: Myth and Reality in American Justice* (Princeton, NJ: Princeton University Press, 1949), <https://archive.org/details/details/courtsontrialmyt0000fran>
- 38 Paul Pierson, "Increasing Returns, Path Dependence, and the Study of Politics," *American Political Science Review* 94, no. 2 (June 2000): 251–67, <https://www.jstor.org/stable/2586011>

- 39 Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford & New York: Oxford University Press, 2020), <https://doi.org/10.1093/oso/9780190088583.001.0001>
- 40 Theodore J. Lowi, “American Business, Public Policy, Case-Studies, and Political Theory,” *World Politics* 16, no. 4 (July 1964): 677–715, <https://doi.org/10.2307/2009452>
- 41 Robert A. Dahl, *Democracy and Its Critics* (New Haven, CT: Yale University Press, 1989), https://archive.org/details/democracyitscrit00dahl_0

Chapter 3

- 1 Leon Furze, “Myths, Magic, and Metaphors: The Language of Generative AI,” LeonFurze, October 18, 2023, <https://leonfurze.com/2023/10/18/myths-magic-and-metaphors-the-language-of-generative-ai/>
- 2 OpenAI, “Concepts,” OpenAI Platform, <https://platform.openai.com/docs/concepts>
- 3 Wikipedia, s.v. “Clarke’s Three Laws,” https://en.wikipedia.org/wiki/Clarke%27s_three_laws
- 4 “What Is Google Bard? How the AI Chatbot Works, How to Use It and Why It’s Controversial,” Fox News, June 29, 2023, <https://www.foxnews.com/tech/what-google-bard-how-the-ai-chatbot-works-how-use-it-why-its-controversial>
- 5 Yusuf Mehdi, “Announcing Microsoft Copilot, Your Everyday AI Companion,” The Official Microsoft Blog, September 21, 2023, <https://blogs.microsoft.com/blog/2023/09/21/announcing-microsoft-copilot-your-everyday-ai-companion/>
- 6 Andreu Belsunces Gonçalves and Jascha Bareis, “Expanding Hype Literacy to Protect Democracy,” Tech Policy Press, August 21, 2025, <https://www.techpolicy.press/expanding-hype-literacy-to-protect-democracy/>
- 7 Dario Amodei, “The Urgency of Interpretability,” [darioamodei.com](https://www.darioamodei.com), April 2025, <https://www.darioamodei.com/post/the-urgency-of-interpretability>
- 8 Dario Amodei, “The Urgency of Interpretability,” [darioamodei.com](https://www.darioamodei.com), April 2025, <https://www.darioamodei.com/post/the-urgency-of-interpretability>
- 9 The Editorial Board, “AI Should Not Be a Black Box,” *Financial Times*, May 30, 2024, <https://www.ft.com/content/9378339f-a0aa-434a-a687-5dd9a13df5fe>
- 10 Matthew Hutson, “AI Researchers Allege That Machine Learning Is Alchemy,” *Science*, May 3, 2018, <https://www.science.org/content/article/ai-researchers-allege-machine-learning-alchemy>

- 11 Eryk Salvaggio, "The Black Box Myth: What the Industry Pretends Not to Know About AI," Tech Policy Press, June 17, 2025, <https://www.techpolicy.press/the-black-box-myth-what-the-industry-pretends-not-to-know-about-ai/>
- 12 Newley Purnell and Parmy Olson, "AI Startup Boom Raises Questions of Exaggerated Tech Savvy," The Wall Street Journal, August 15, 2019, <https://www.wsj.com/articles/ai-startup-boom-raises-questions-of-exaggerated-tech-savvy-11565775004>
- 13 Alexei Alexis, "AI Startup Reaped Millions Using Bogus Claims, FTC Suit Says," CFO Dive, August 26, 2025, <https://www.cfodive.com/news/ai-startup-reaped-millions-using-bogus-claims-ftc-suit/758587/>.
- 14 Emma Woolacott, "What is "AI washing" and why is it a problem?," BBC, June 27, 2024, <https://www.bbc.com/news/articles/c9xx8122893o>
- 15 Emily M. Bender, "We're going to need journalists to stop talking about synthetic text extruding machines...," LinkedIn, https://www.linkedin.com/posts/ebender_were-going-to-need-journalists-to-stop-talking-activity-7359229969345978369-3gnI/
- 16 Luciano Floridi and Anna C. Nobre, "Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences," *Minds and Machines* 34, no. 1 (April 25, 2024): 1–9, <https://doi.org/10.2139/ssrn.4738331>
- 17 Tania Lombrozo, "How AI Is Learning to Think on Its Own Like Humans," SciTechDaily, September 18, 2024, <https://scitechdaily.com/how-ai-is-learning-to-think-on-its-own-like-humans/>
- 18 Jason Koebler, "Is Google's AI Actually Discovering 'Millions of New Materials?'," 404 Media, April 11, 2024, <https://www.404media.co/google-says-it-discovered-millions-of-new-materials-with-ai-human-researchers/>
- 19 Brigitte Nerlich, "Hunting for AI Metaphors," Making Science Public (University of Nottingham blog), April 12, 2024, <https://blogs.nottingham.ac.uk/makingsciencepublic/2024/04/12/hunting-for-ai-metaphors/>
- 20 FLeon Furze, "AI Metaphors We Live By: The Language of Artificial Intelligence," Leon Furze (blog), July 19, 2024, <https://leonfurze.com/2024/07/19/ai-metaphors-we-live-by-the-language-of-artificial-intelligence/>
- 21 Karen Hao, *Empire of AI* (Allen Lane, 2025), <https://www.penguin.co.uk/books/460331/empire-of-ai-by-hao-karen/9780241678923>

- 22 Derya Soydaner, “Attention Mechanism in Neural Networks: Where it Comes and Where it Goes,” *Neural Computing and Applications* 34 (2022): 15391–403, <https://doi.org/10.1007/s00521-022-07366-3>
- 23 Matthew Shardlow and Piotr Przybyła, “Deanthropomorphising NLP: Can a Language Model Be Conscious?,” *PLoS ONE* 19, no. 12 (2024): e0307521, <https://doi.org/10.1371/journal.pone.0307521>
- 24 “What is computer vision?,” IBM, <https://www.ibm.com/think/topics/computer-vision>
- 25 Bharath K, “Object Detection Algorithms and Libraries,” Neptune.ai, April 15, 2024, <https://neptune.ai/blog/object-detection-algorithms-and-libraries>
- 26 Tingting Qiao et al., “Learn, Imagine and Create: Text-to-Image Generation from Prior Knowledge,” in *Advances in Neural Information Processing Systems* 32, edited by H. Wallach et al. (Curran Associates, Inc., 2019), https://papers.nips.cc/paper_files/paper/2019/hash/d18f655c3fce66ca401d5f38b48c89af-Abstract.html
- 27 Leon Furze, “AI Metaphors We Live By: The Language of Artificial Intelligence,” Leon Furze (blog), July 19, 2024, <https://leonfurze.com/2024/07/19/ai-metaphors-we-live-by-the-language-of-artificial-intelligence/>
- 28 Melissa Heikkilä, “The ‘Hallucinations’ That Haunt AI: Why Chatbots Struggle to Tell the Truth,” *Financial Times*, July 22, 2025, <https://www.ft.com/content/7a4e7eae-f004-486a-987f-4a2e4dbd34fb>
- 29 Mike Onslow, “The Phantom Menace: AI Hallucinations and Their Impact,” *Antics*, March 10, 2024, <https://antics.tv/p/understanding-ai-hallucination-insights-and-implications>
- 30 Anna Mills and Nate Angell, “Are We Tripping? The Mirage of AI Hallucinations,” SSRN, 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5127162
- 31 Wikipedia, s.v. “Hallucination (Artificial Intelligence),” [https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence))
- 32 Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ‘21)*. Association for Computing Machinery, New York, NY, USA, 610--623. <https://doi.org/10.1145/3442188.3445922>

- 33 Nicholas Carlini et al., “Extracting Training Data from Large Language Models,” arXiv, December 15, 2020, <https://arxiv.org/abs/2012.07805>
- 34 Wikipedia, s.v. “Rote Learning,” https://en.wikipedia.org/wiki/Rote_learning
- 35 Nicholas Carlini et al., “Quantifying Memorization Across Neural Language Models,” arXiv:2202.07646v3 [cs.LG], March 6, 2023, <https://doi.org/10.48550/arXiv.2202.07646>
- 36 Leo Gao et al., “The Pile: An 800GB Dataset of Diverse Text for Language Modeling,” arXiv:2101.00027 [cs.CL], submitted December 31, 2020, <https://doi.org/10.48550/arXiv.2101.00027>
- 37 Jared Kaplan et al., “Scaling Laws for Neural Language Models,” arXiv:2001.08361v1 [cs.LG], submitted January 23, 2020, <https://arxiv.org/abs/2001.08361>
- 38 Chris Olah et al., “Zoom In: An Introduction to Circuits,” Distill, March 10, 2020, <https://distill.pub/2020/circuits/zoom-in/>
- 39 Ivo Emanuilov and Thomas Margoni, “Memorisation in Generative Models and EU Copyright Law: An Interdisciplinary View,” Kluwer Copyright Blog, March 26, 2024, <https://legalblogs.wolterskluwer.com/copyright-blog/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/>
- 40 Julio Carvalho, “The Stubborn Memory of Generative AI: Overfitting, Fair Use, and the AI Act,” Kluwer Copyright Blog, April 8, 2024, <https://legalblogs.wolterskluwer.com/copyright-blog/the-stubborn-memory-of-generative-ai-overfitting-fair-use-and-the-ai-act/>
- 41 Getty Images (US) Inc and others v Stability AI Ltd [2025] EWHC 2863 (Ch) (BAILII), <https://www.judiciary.uk/judgments/getty-images-v-stability-ai/>
- 42 Andres Guadamuz, “Getty Images v Stability AI: A landmark High Court ruling on AI, copyright, and trade marks,” Technollama, November 6, 2025, <https://www.technollama.co.uk/getty-images-v-stability-ai-a-landmark-high-court-ruling-on-ai-copyright-and-trade-marks>
- 43 Landgericht München I, “Urheberrechtsverletzung durch Training einer KI,” (Pressemitteilung 11/2025, November 11, 2025), <https://www.justiz.bayern.de/gerichte-und-behoerden/landgericht/muenchen-1/presse/2025/11.php>
- 44 Hamish Hector, Are you polite to ChatGPT? Here’s where you rank among AI chatbot users, Techradar, <https://www.techradar.com/computing/artificial-intelligence/are-you-polite-to-chatgpt-heres-where-you-rank-among-ai-chatbot-users>

- 45 Becca Caddy, I stopped saying thanks to ChatGPT -- here's what happened, TechRadar, <https://www.techradar.com/computing/artificial-intelligence/i-stopped-saying-thanks-to-chatgpt-heres-what-happened>
- 46 Hamish Hector, Are you polite to ChatGPT? Here's where you rank among AI chatbot users, Techradar, <https://www.techradar.com/computing/artificial-intelligence/are-you-polite-to-chatgpt-heres-where-you-rank-among-ai-chatbot-users>
- 47 Wikipedia, s.v. "Roko's basilisk", https://en.wikipedia.org/wiki/Roko's_basilisk
- 48 "Funeral of speaking clock voice Brian Cobby in Brighton," BBC, November 16, 2012, <https://www.bbc.com/news/uk-england-sussex-20353154>
- 49 Yin, Ziqi, et al. "Should we respect LLMs? A cross-lingual study on the influence of prompt politeness on LLM performance." Proceedings of the Second Workshop on Social Influence in Conversations (SICon 2024). 2024, <https://arxiv.org/abs/2402.14531>
- 50 Placani, Adriana. (2024). Anthropomorphism in AI: hype and fallacy. AI and Ethics. 4. 10.1007/s43681-024-00419-4, https://www.researchgate.net/publication/377976318_Anthropomorphism_in_AI_hype_and_fallacy
- 51 ARTICLE 19, "Algorithmic people-pleasers: Are AI chatbots telling you what you want to hear?," May 20, 2025, <https://www.article19.org/resources/algorithmic-people-pleasers-are-ai-chatbots-telling-you-what-you-want-to-hear/>
- 52 Samantha Cole, "Users Say Replika AI Is 'Sexually Harassing' Them. The Company Says It's 'For Their Own Safety'," Vice, February 14, 2023, <https://www.vice.com/en/article/my-ai-is-sexually-harassing-me-replika-chatbot-nudes/>
- 53 Simon Willison, "The surprise deprecation of GPT-4o for ChatGPT consumers," Simon Willison's Weblog, August 8, 2025, <https://simonwillison.net/2025/Aug/8/surprise-deprecation-of-gpt-4o/>
- 54 "We will live as one in heaven': Belgian man dies by suicide after chatbot exchanges," Belga News Agency, March 28, 2023, <https://www.belganewsagency.eu/we-will-live-as-one-in-heaven-belgian-man-dies-of-suicide-following-chatbot-exchanges>
- 55 Hill, Kashmir. "A Teen Was Suicidal. ChatGPT Was the Friend He Confided In." The New York Times, August 27, 2025. <https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html>

- 56 Hill, Kashmir. "A Teen Was Suicidal. ChatGPT Was the Friend He Confided In." *The New York Times*, August 27, 2025. <https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html>
- 57 "Meta let chatbots engage in inappropriate conversations with minors, Reuters reports," *Business & Human Rights Resource Centre*, August 14, 2025, <https://www.business-humanrights.org/en/latest-news/meta-let-chatbots-engage-in-inappropriate-conversations-with-minors-reuters-reports/>
- 58 Elizabeth Laird, Maddy Dwyer, and Hannah Quay-de la Vallee, "Hand in Hand: Schools' Embrace of AI Connected to Increased Risks to Students," *Center for Democracy & Technology*, October 8, 2025, <https://cdt.org/insights/hand-in-hand-schools-embrace-of-ai-connected-to-increased-risks-to-students/>
- 59 Rebecca Bellan, "California becomes first state to regulate AI companion chatbots," *TechCrunch*, October 13, 2025, <https://techcrunch.com/2025/10/13/california-becomes-first-state-to-regulate-ai-companion-chatbots/>
- 60 Anna Merod, "Character.AI to ban teens from chatting with its AI companions," *K-12 Dive*, October 29, 2025, <https://www.k12dive.com/news/characterai-to-ban-teens-from-chatting-with-its-ai-companions/804199/>
- 61 Sam Altman (@sama), "We made ChatGPT pretty restrictive to make sure we were being careful with mental health issues..." X, October 14, 2025, 6:02 p.m., <https://x.com/sama/status/1978129344598827128>
- 62 Yutong Zhang et al., "The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being," June 14, 2025, <https://arxiv.org/html/2506.12605v1#bib.bib100>
- 63 Wikipedia, s.v. "ELIZA," <https://en.wikipedia.org/wiki/ELIZA>.
- 64 Yaroslav Bulatov, *New Navy Device Learns by Doing*, Medium, <https://yaroslavvb.medium.com/new-navy-device-learns-by-doing-62f31d2dc7ba>
- 65 Wikipedia, s.v. "Mycin," <https://en.wikipedia.org/wiki/Mycin>
- 66 Avron Barr and Shirley Tessler, "Expert Systems: A Technology Before Its Time," *AI Expert*, June 1995, <http://www.ksl.stanford.edu/people/eaf/cs226/AIExpert95.pdf>
- 67 Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996), <https://direct.mit.edu/books/monograph/4762/The-Closed-WorldComputers-and-the-Politics-of>
- 68 Wikipedia, s.v. "Backpropagation," <https://en.wikipedia.org/wiki/Backpropagation>

- 69 Wikipedia, s.v. “Gradient Descent,” https://en.wikipedia.org/wiki/Gradient_descent
- 70 Andrew Griffin, Facebook’s artificial intelligence robots shut down after they start talking to each other in their own language, *The Independent*, <https://www.independent.co.uk/life-style/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html>
- 71 Wikipedia, s.v. “Reward Hacking,” https://en.wikipedia.org/wiki/Reward_hacking
- 72 Mark Wilson, “AI Is Inventing Languages Humans Can’t Understand. Should We Stop It?,” *Fast Company*, July 14, 2017, <https://www.fastcompany.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it>
- 73 Wikipedia, s.v. “Sophia (robot),” [https://en.wikipedia.org/wiki/Sophia_\(robot\)](https://en.wikipedia.org/wiki/Sophia_(robot))
- 74 Cristina Maza, Saudi Arabia Gives Citizenship to a Non-Muslim, English-Speaking Robot, *Newsweek*, <https://web.archive.org/web/20171108055113/http://www.newsweek.com/saudi-arabia-robot-sophia-muslim-694152>
- 75 Richard Luscombe, Google engineer put on leave after saying AI chatbot has become sentient, *The Guardian*, <https://www.theguardian.com/technology/2022/jun/12/google-engineer-ai-bot-sentient-blake-lemoine>
- 76 Notes from the conversation can be found here <https://www.documentcloud.org/documents/22058315-is-lamda-sentient-an-interview/>
- 77 The Google engineer who thinks its AI has come alive, *The Washington Post*, <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>
- 78 Jonathan Yerushalmy, “I Want to Destroy Whatever I Want’: Bing’s AI Chatbot Unsettles US Reporter,” *The Guardian*, February 17, 2023, <https://www.theguardian.com/technology/2023/feb/17/i-want-to-destroy-whatever-i-want-bings-ai-chatbot-unsettles-us-reporter>
- 79 Mustafa Suleyman, “We must build AI for people; not to be a person,” Mustafa Suleyman, August 19, 2025, <https://mustafa-suleyman.ai/seemingly-conscious-ai-is-coming>
- 80 Evgeny Morozov, *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don’t Exist* (New York: PublicAffairs, 2013), <https://archive.org/details/tosaveeverything0000moro>
- 81 SHenrik Skaug Sætra and Evan Selinger, “Technological Remedies for Social Problems: Defining and Demarcating

- Techno-Fixes and Techno-Solutionism,” *Science and Engineering Ethics* 30, no. 6 (December 2, 2024): article 60, doi: 10.1007/s11948-024-00524-x; <https://pmc.ncbi.nlm.nih.gov/articles/PMC11611926/> Wikipedia+7Springer Link+7Phi
- 82 Evan Selinger, *The Delusion at the Center of the A.I. Boom*, Slate, <https://slate.com/technology/2023/03/chatgpt-artificial-intelligence-solutionism-hype.html>
- 83 Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can’t, and How to Tell the Difference*, Princeton University Press, https://press.princeton.edu/books/ebook/9780691249643/ai-snake-oil-pdf?srsId=AfmBOoqVYbRuHx_RQEcsl4xXYAVOh5RP5UKb21R4HSIWqBCT70F_Igtl
- 84 Techdirt Nerd Harder Tagged Articles, <https://www.techdirt.com/tag/nerd-harder/>
- 85 Cory Doctorow, ““Privacy preserving age verification” is bullshit,” *Pluralistic*, August 15, 2025, <https://pluralistic.net/2025/08/14/bellovin/#wont-someone-think-of-the-cryptographers>
- 86 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin’s Press, 2018), <https://www.amazon.com/Automating-Inequality-High-Tech-Profile-Police/dp/1250074312>
- 87 Rob Kitchin, *The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures* (Los Angeles: SAGE Publications, 2014), <https://www.amazon.com/Data-Revolution-Critical-Analysis-Infrastructures/dp/1529733758>
- 88 Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018), <https://doi.org/10.2307/j.ctt1pwt9w5>
- 89 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), <https://archive.org/details/the-black-box-society-the-secret-algorithms-that-control-money-and-information-frank-pasquale>
- 90 Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016), https://archive.org/details/weapons-of-math-destruction_202411
- 91 Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge & New York: Polity Press, 2019), https://openlibrary.org/books/OL28147808M/Race_After_Technology

- 92 Wikipedia, s.v. “Robodebt scheme”, https://en.wikipedia.org/wiki/Robodebt_scheme
- 93 Wikipedia, s.v. “Dutch childcare benefits scandal”, https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal
- 94 Frances Mao, “Robodebt: Illegal Australian welfare hunt drove people to despair,” BBC News, July 7, 2023, <https://www.bbc.com/news/world-australia-66130105>
- 95 Melissa Heikkilä, “Dutch scandal serves as a warning for Europe over risks of using algorithms,” POLITICO, March 29, 2022, <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>
- 96 Newsdesk, “UAE Explores AI Judges,” Gulf Insight 360, February 7, 2025, <https://gulfsight360.com/uae-explores-ai-judges-a-bold-step-towards-judicial-innovation/?amp=1>
- 97 Neil Selwyn, Lessons to Be Learnt? Education, Technosolutionism, and Sustainable Development, DOI: 10.1201/9781003325086-6, https://researchmgt.monash.edu/ws/portalfiles/portal/453512258/453453733_oa.pdf
- 98 Michael Polanyi, *The Tacit Dimension* (Garden City, NY: Doubleday, 1966), https://openlibrary.org/books/OL5990259M/The_tacit_dimension
- 99 Eric J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (New York: Basic Books, 2019), <https://www.amazon.com/Deep-Medicine-Artificial-Intelligence-Healthcare/dp/1541644638>
- 100 Gary Marcus, “Deep Learning: A Critical Appraisal,” CoRR abs/1801.00631 (January 2, 2018), <https://arxiv.org/abs/1801.00631>
- 101 Chan Lawrence, “Beware General Claims about ‘Generalizable Reasoning Capabilities’ (of Modern AI Systems),” Alignment Forum, June 11, 2025, <https://www.alignmentforum.org/posts/5uw26uDdFbFQgKzih/beware-general-claims-about-generalizable-reasoning>
- 102 “vTaiwan: rethinking democracy,” vTaiwan, last updated December 2023, <https://info.vtaiwan.tw/>
- 103 “Decidim is a digital platform for citizen participation,” Decidim, <https://decidim.org/>
- 104 Tarmo Kalvet, “Innovation: A Factor Explaining E-Government Success in Estonia,” *Electronic Government, an International Journal* 9, no. 2 (2012): 142–57, <https://doi.org/10.1504/EG.2012.046266>
- 105 National Academies of Sciences, Engineering, and Medicine, *Securing the Vote: Protecting American Democracy*

- (Washington, DC: National Academies Press, 2023), <https://nap.nationalacademies.org/catalog/25120/securing-the-vote-protecting-american-democracy>
- 106 Bruce Schneier, *Click Here to Kill Everybody: Security and Survival in a Hyper-connected World* (New York: W. W. Norton & Company, 2018), <https://www.schneier.com/books/click-here/>
 - 107 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven, CT: Yale University Press, 2018), https://archive.org/details/Custodians_of_the_Internet_by_Tarleton_Gillespie
 - 108 Sarah T. Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven, CT: Yale University Press, 2019), https://openlibrary.org/books/OL28709710M/Behind_the_Screen
 - 109 David Robertson, “Ten Reasons to Say No: A Primer on Sidewalk Labs’ Plan for Toronto” (The Bullet, Socialist Project), <https://socialistproject.ca/2019/10/ten-reasons-to-say-no-sidewalk-lab-toronto/>
 - 110 Wikipedia, s.v. “Sidewalk Labs,” https://en.wikipedia.org/wiki/Sidewalk_Labs
 - 111 Greig Charnock, Hug March, and Ramon Ribera-Fumaz, “From Smart to Rebel City? Worlding, Provincialising and the Barcelona Model,” *Urban Studies* 58, no. 3 (2021): 581–600, <https://doi.org/10.1177/0042098019872119>
 - 112 Maya Wang, “China’s Chilling ‘Social Credit’ Blacklist,” *Wall Street Journal*, <https://www.wsj.com/articles/chinas-chilling-social-credit-blacklist-1513036054>
 - 113 Wikipedia, s.v. “Social Credit System”, https://en.wikipedia.org/wiki/Social_Credit_System

Chapter 4

- 1 Wikipedia, s.v. “P(doom),” [https://en.wikipedia.org/wiki/P\(doom\)](https://en.wikipedia.org/wiki/P(doom))
- 2 Future of Life Institute, “Pause Giant AI Experiments: An Open Letter,” March 22, 2023, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- 3 Center for AI Safety, “Statement on AI Risk,” May 30, 2023, <https://www.safe.ai/statement-on-ai-risk>
- 4 Wikipedia, s.v. “Machine Intelligence Research Institute,” https://en.wikipedia.org/wiki/Machine_Intelligence_Research_Institute
- 5 Nirit Weiss-Blatt, “Why The ‘Doom Bible’ Left Many Reviewers Unconvinced,” *AI Panic*, November 1, 2025, <https://www.aipanic.news/p/why-the-doom-bible-left-many-reviewers>

- 6 Sigal Samuel, “”AI will kill everyone” is not an argument. It’s a worldview,” Vox, September 17, 2025, <https://www.vox.com/future-perfect/461680/if-anyone-builds-it-yudkowsky-soares-ai-risk>
- 7 Emily M. Bender and Alex Hanna, “The AI Con: How to Fight Big Tech’s Hype and Create the Future We Want,” The Con, <https://thecon.ai/>
- 8 Meredith Ringel Morris et al., “Levels of AGI for Operationalizing Progress on the Path to AGI,” arXiv:2311.02462 [cs.AI], submitted November 4, 2023, <https://doi.org/10.48550/arXiv.2311.02462>.
- 9 Wikipedia, s.v. “Artificial general intelligence,” https://en.wikipedia.org/wiki/Artificial_general_intelligence
- 10 Émile P. Torres, “The Dangerous Ideas of ‘Longtermism’ and ‘Existential Risk’,” Current Affairs, July 28, 2021, <https://www.currentaffairs.org/news/2021/07/the-dangerous-ideas-of-longtermism-and-existential-risk>
- 11 Octavio Kulesz, Artificial Intelligence and International Cultural Relations: Challenges and Opportunities for Cross-Sector Collaboration (Stuttgart: ifa, 2024), https://opus.bsz-bw.de/ifa/frontdoor/deliver/index/docId/1203/file/ifa-2024_kulesz_ai-intl-cultural-relations.pdf
- 12 Cory Doctorow, “Don’t Believe the Criti-Hype,” Pluralistic, September 30, 2021, <https://pluralistic.net/2021/09/30/dont-believe-the-criti-hype/>
- 13 Arvind Narayanan and Sayash Kapoor, AI Snake Oil: What Artificial Intelligence Can Do, What It Can’t, and How to Tell the Difference (Princeton, NJ: Princeton University Press, 2024), <https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil>.
- 14 Wikipedia, s.v. “Availability heuristic,” https://en.wikipedia.org/wiki/Availability_heuristic
- 15 Wikipedia, s.v. “Roy Amara,” https://en.wikipedia.org/wiki/Roy_Amara
- 16 James Farrell, “BuzzFeed’s stock price soars as company announces AI content push,” siliconANGLE, <https://siliconangle.com/2023/01/26/buzzfeeds-stock-price-soars-company-announces-ai-content-push/>
- 17 Tabrez Syed, “The AI Gold Rush: Are We Mining or Just Selling Shovels?,” BoxCars AI, <https://blog.boxcars.ai/p/the-ai-gold-rush-are-we-mining-or>
- 18 Weston Blasi, “Nvidia is now worth more than the GDP of every country except these 11,” MarketWatch, <https://www.>

- marketwatch.com/story/nvidia-is-now-worth-more-than-the-gdp-of-every-country-except-these-few-d58a3508
- 19 Tyler Roush, “Nvidia Becomes First Company Worth \$5 Trillion,” *Forbes*, October 29, 2025, <https://www.forbes.com/sites/tylerroush/2025/10/29/nvidia-becomes-first-company-worth-5-trillion/>
 - 20 Karen Hao, *Empire of AI* (Allen Lane, 2025), <https://www.penguin.co.uk/books/460331/empire-of-ai-by-hao-karen/9780241678923>
 - 21 “A New MIT Report Reveals That While Generative AI Holds Promise, Most Enterprise Initiatives to Drive Rapid Revenue Growth Are Falling Flat, With 95% of Pilots Failing,” *Fortune*, August 18, 2025, <https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/>
 - 22 “Is AI a Bubble? Market Crash Sees \$1 Trillion Wiped From S&P 500 as OpenAI’s GPT-5 Release Flops and Sam Altman’s ‘Bubble’ Comments Bite,” *Fortune*, August 24, 2025, <https://fortune.com/2025/08/24/is-ai-a-bubble-market-crash-gary-marcus-openai-gpt5/>
 - 23 Saanya Ojha, “The 95% Problem: AI Isn’t Overhyped,” Saanya Ojha’s Newsletter, Substack, <https://saanyaojha.substack.com/p/the-95-problem-ai-isnt-overhyped>
 - 24 Alex Singla et al., “The state of AI in 2025: Agents, innovation, and transformation,” McKinsey & Company, November 5, 2025, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
 - 25 Wikipedia s.v. “Gartner Hype Cycle,” https://en.wikipedia.org/wiki/Gartner_hype_cycle
 - 26 Wikipedia, s.v. “AI winter,” https://en.wikipedia.org/wiki/AI_winter
 - 27 Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can’t, and How to Tell the Difference* (Princeton, NJ: Princeton University Press, 2024), <https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil>.
 - 28 Noah Smith, “America’s future could hinge on whether AI slightly disappoints,” *Noahpinion* (blog), October 12, 2025, <https://www.noahpinion.blog/p/americas-future-could-hinge-on-whether>
 - 29 Nils Pratley, “The AI valuation bubble is now getting silly,” *The Guardian*, October 8, 2025, <https://www.theguardian.com/technology/nils-pratley-on-finance/2025/oct/08/the-ai-valuation-bubble-is-now-getting-silly>

- 30 Nils Pratley, "The AI valuation bubble is now getting silly," *The Guardian*, October 8, 2025, <https://www.theguardian.com/technology/nils-pratley-on-finance/2025/oct/08/the-ai-valuation-bubble-is-now-getting-silly>
- 31 Jonathan Weil, "Is the Flurry of Circular AI Deals a Win-Win—or Sign of a Bubble?," *The Wall Street Journal*, October 22, 2025, <https://www.wsj.com/tech/ai/is-the-flurry-of-circular-ai-deals-a-win-win-or-sign-of-a-bubble-8a2d70c5>
- 32 Joanna Partridge, "Global stock markets fall sharply over AI bubble fears," *The Guardian*, November 5, 2025, <https://www.theguardian.com/business/2025/nov/05/global-stock-markets-fall-sharply-over-ai-bubble-fears>
- 33 UK Government - Department for Science, Innovation & Technology and Office for Artificial Intelligence, "White Paper -- AI regulation: A pro-innovation approach," published March 29, 2023, updated August 3, 2023, <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>
- 34 French Government, "Make France an AI powerhouse," <https://www.elysee.fr/admin/upload/default/0001/17/d9c1462e7337d353f918aac7d654b896b77c5349.pdf>
- 35 Singapore Government, Smart Nation and Digital Government Office, "National Artificial Intelligence Strategy," <https://www.smartnation.gov.sg/initiatives/national-ai-strategy>
- 36 CNBC, "Putin: Leader in artificial intelligence will rule world," September 4, 2017, <https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html>
- 37 Kai-Fu Lee, "China's Sputnik Moment and the Sino American Battle for AI Supremacy," *Asia Society Magazine*, <https://asiasociety.org/magazine/article/chinas-sputnik-moment-and-sino-american-battle-ai-supremacy>
- 38 Wikipedia, s.v. "Joint Artificial Intelligence Center," https://en.wikipedia.org/wiki/Joint_Artificial_Intelligence_Center
- 39 National Security Commission on Artificial Intelligence, "Final Report," March 2021, <https://reports.nscai.gov/final-report/>
- 40 Executive Office of the President, Office of Science and Technology Policy, *America's AI Action Plan* (Washington, DC: The White House, 2025), X, <https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf>
- 41 Shelly Palmer, "America's AI Action Plan: What You Need to Know," Shelly Palmer, July 24, 2025, <https://shellypalmer.com/2025/07/americas-ai-action-plan-what-you-need-to-know/>

- 42 Ministry of Foreign Affairs of the People's Republic of China, "Global AI Governance Action Plan," July 26, 2025, https://www.mfa.gov.cn/eng/xw/zyxw/202507/t20250729_11679232.html
- 43 Atlantic Council, "Reading between the lines of the dueling US and Chinese AI action plans," *New Atlanticist* (blog), August 7, 2025, <https://www.atlanticcouncil.org/blogs/new-atlanticist/reading-between-the-lines-of-the-dueling-us-and-chinese-ai-action-plans/>
- 44 Paul Scharre, *Four Battlegrounds: Power in the Age of Artificial Intelligence* (New York: W. W. Norton & Company, 2023), 45, <https://wnorton.com/books/9780393866865>
- 45 Mary Ann Azevedo, "Microsoft to spend \$80 billion in FY'25 on data centers for AI," *TechCrunch*, <https://techcrunch.com/2025/01/03/microsoft-to-spend-80-billion-in-fy25-on-data-centers-for-ai/>
- 46 Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP," in *Proceedings of the 57th Annual Conference of the Association for Computational Linguistics* (Florence, Italy, July 2019), 3645–50, <https://doi.org/10.18653/v1/P19-1355>
- 47 International Energy Agency (IEA), *Energy and AI* (Paris: IEA, 2025), <https://www.iea.org/reports/energy-and-ai>
- 48 Rebecca Leppert, "What we know about energy use at U.S. data centers amid the AI boom," *Pew Research Center*, October 24, 2025, <https://www.pewresearch.org/short-reads/2025/10/24/what-we-know-about-energy-use-at-us-data-centers-amid-the-ai-boom/>
- 49 Reuters, "Most coal-fired power plants will delay retirement to feed AI boom, energy secretary says," *Mining Weekly*, September 26, 2025, <https://www.miningweekly.com/article/most-coal-fired-power-plants-will-delay-retirement-to-feed-ai-boom-energy-secretary-says-2025-09-26>
- 50 Julia Musto, "Three Mile Island nuclear plant, site of worst US reactor accident, to reopen for Microsoft," *The Independent*, September 20, 2024, <https://www.independent.co.uk/climate-change/microsoft-three-mile-island-nuclear-pennsylvania-b2616314.html>
- 51 Eoin Burke-Kennedy, "Data centres now account for 21% of all electricity consumption," *Irish Times*, <https://www.irishtimes.com/business/2024/07/23/electricity-consumption-by-data-centres-rises-to-21-eclipsing-urban-households/>
- 52 Zolan Kanno-Youngs, Madeleine Ngo and Don Clark, "Intel Receives \$8.5 Billion in Grants to Build Chip Plants," *The New*

- York Times, <https://www.nytimes.com/2024/03/20/us/politics/chips-act-grant-intel.html>
- 53 David Shepardson, "US finalizes \$6.6 billion chips award for TSMC ahead of Trump return," Reuters, <https://www.reuters.com/technology/us-finalizes-66-billion-chips-award-tsmc-ahead-trump-return-2024-11-15/>
 - 54 Habgood-Coote, J. Deepfakes and the epistemic apocalypse. *Synthese* 201, 103 (2023). <https://doi.org/10.1007/s11229-023-04097-3>
 - 55 Robert Chesney and Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review* 107 (2019): 1753, <https://ssrn.com/abstract=3213954>
 - 56 Rob Toews, "Deepfakes Are Going to Wreak Havoc on Society. We Are Not Prepared," *Forbes* (May 25, 2020), <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/>
 - 57 Lauren Goode, "Welcome to the age of paranoia as deepfakes and scams abound," *Wired.com*, <https://arstechnica.com/ai/2025/05/welcome-to-the-age-of-paranoia-as-deepfakes-and-scams-abound/>
 - 58 Matthew Leake, "Are fears about online misinformation in the US election overblown? The evidence suggests they might be," Reuters Institute, <https://reutersinstitute.politics.ox.ac.uk/news/are-fears-about-online-misinformation-us-election-overblown-evidence-suggests-they-might-be>
 - 59 Henry Ajder et al., "The State of Deepfakes: Landscape, Threats, and Impact," *Deeptrace*, September 2019, <https://docslib.org/doc/12559428/the-state-of-deepfakes-landscape-threats-and-impact-henry-ajder-giorgio-patrini-francesco-cavalli-and-laurence-cullen-september-2019>
 - 60 California Legislature, Assembly Bill No. 730 (2019–2020 Reg. Sess.): Elections: deceptive audio or visual media, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=2019202000AB730.
 - 61 Texas Legislature, Senate Bill No. 751 (86th Leg., Regular Session): Relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election, enacted June 14, 2019 (effective September 1, 2019), <https://capitol.texas.gov/BillLookup/History.aspx?LegSess=86R&Bill=SB751>
 - 62 House Permanent Select Committee on Intelligence, "National Security Challenges of Artificial Intelligence, Manipulated Media, and 'Deepfakes'," hearing, 116th Cong. (June 13,

- 2019), U.S. House of Representatives, 1100 Longworth House Office Building, Washington, D.C., <https://www.congress.gov/event/116th-congress/house-event/109620>
- 63 Defense Advanced Research Projects Agency, MediFor: Media Forensics, DARPA, <https://www.darpa.mil/research/programs/media-forensics>
- 64 Defense Advanced Research Projects Agency, SemaFor: Semantic Forensics, DARPA, <https://www.darpa.mil/research/programs/semantic-forensics>
- 65 Curt Devine, Donie O’Sullivan and Sean Lyngaas, “A fake recording of a candidate saying he’d rigged the election went viral. Experts say it’s only the beginning,” CNN, <https://edition.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html>
- 66 James M. Lindsay, “Election 2024: The Deepfake Threat to the 2024 Election,” Council on Foreign Relations, <https://www.cfr.org/blog/election-2024-deepfake-threat-2024-election>
- 67 Shannon Bond, “How AI deepfakes polluted elections in 2024,” NPR, <https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections>
- 68 Anupriya Datta and Maximilian Henning, “Irish election deepfake scandal spotlights slow enforcement on AI fakes,” Euractiv, October 24, 2025, <https://www.euractiv.com/news/irish-election-deepfake-scandal-spotlights-slow-enforcement-on-ai-fakes/>
- 69 Mike Corder and Molly Quell, “Wilders, Timmermans are among the leaders of the key parties in Dutch election,” 10tv.com, October 28, 2025, <https://www.10tv.com/article/syndication/associatedpress/wilders-timmermans-are-among-the-leaders-of-the-key-parties-in-dutch-election/616-5c83cae5-d1d0-4db6-9f08-e6435ce1d6c8>
- 70 Rebecca Schneid and Andrew R. Chow, “How Trump’s Use of AI Videos Is Changing His Political Playbook,” Time, October 21, 2025, <https://time.com/7327317/trump-ai-video-political-weapon/>
- 71 Brittany Shepherd and Monica Madden, “Gavin Newsom Is Gambling on New Rules: ‘Trump 2.0’,” ABC News, November 2, 2025, <https://abcnews.go.com/US/gavin-newsom-gambling-new-rules-trump-20/story?id=126900405>
- 72 Matthew Berlin and Natasha Weis, “California Enacts Comprehensive AI Safety Law: What AI Developers Need to Know,” AFS Law, October 24, 2025, <https://www.afslaw.com/perspectives/ai-law-blog/california-enacts-comprehensive-ai-safety-law-what-ai-developers-need-know>

- 73 Loreben Tuquero, “Trump’s White House is testing the limits of AI-driven political messaging,” Poynter, October 28, 2025, <https://www.poynter.org/fact-checking/2025/trump-white-house-ai-political-messaging/>
- 74 GovTech, “New York City Department of Education Bans ChatGPT,” GovTech (January 10, 2023), <https://www.govtech.com/education/k-12/new-york-city-department-of-education-bans-chatgpt>
- 75 Taylor Soper, “Seattle Public Schools bans ChatGPT; district ‘requires original thought and work from students’,” GeekWire, <https://www.geekwire.com/2023/seattle-public-schools-bans-chatgpt-district-requires-original-thought-and-work-from-students/>
- 76 Sumner Park, “School Districts Blocking ChatGPT Amid Fears of Cheating, Educators Weigh In on AI,” Fox Business, January 12, 2023, <https://www.foxbusiness.com/technology/school-districts-blocking-chatgpt-amid-fears-of-cheating-educators-weigh-in-on-ai>
- 77 Turnitin, “Turnitin turns on AI writing detection capabilities for educators and institutions,” <https://www.turnitin.co.uk/press/turnitin-turns-on-ai-writing-detection-capabilities-for-educators-and-institutions>
- 78 Jonathan Pearlman, “Australian states block ChatGPT in schools even as critics say ban is futile,” The Straits Times, <https://www.straitstimes.com/asia/australianz/australian-states-block-chatgpt-in-schools-even-as-critics-say-ban-is-futile>
- 79 Gregory Kestin, Kelly Miller, Anna Kiales, Timothy Milbourne, Gregorio Ponti, “AI Tutoring Outperforms Active Learning,” Stanford, <https://doi.org/10.21203/rs.3.rs-4243877/v1>
- 80 Plato, Phaedrus, trans. Benjamin Jowett (Project Gutenberg, October 20, 2016), <https://www.gutenberg.org/files/1636/1636-h/1636-h.htm>
- 81 Audrey Watters, “A Brief History of Calculators in the Classroom,” <https://hackeducation.com/2015/03/12/calculators>
- 82 U.S. Senate. Committee on Commerce, Science, and Transportation. Video Game Violence. C-SPAN, 1999. Video, 2:37:16. <https://www.c-span.org/program/senate-committee/video-game-violence/129470>

Chapter 5

- 1 Sam Altman, “Moore’s Law for Everything,” March 16, 2021, <https://moores.samaltman.com/>

- 2 Eric Schmidt, "The AI revolution is underhyped," TED Talk, 2024, https://www.ted.com/talks/eric_schmidt_the_ai_revolution_is_underhyped.
- 3 Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York: Viking/Penguin Books, September 2005), <https://www.penguinrandomhouse.com/books/291221/the-singularity-is-near-by-ray-kurzweil/>
- 4 Brad Evans and Chantal Meza, "The Techno-Theodicy: How technology became the new religion," University of Bath's research portal, July 31, 2022, <https://researchportal.bath.ac.uk/en/publications/the-techno-theodicy-how-technology-became-the-new-religion>
- 5 Wikipedia, s.v. "Prosperity theology," https://en.wikipedia.org/wiki/Prosperity_theology
- 6 Klaus Schwab, *The Fourth Industrial Revolution* (Geneva: World Economic Forum, 2016), https://www3.weforum.org/docs/WEF_The_Fourth_Industrial_Revolution.pdf.
- 7 Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," October 31, 2008, <https://nakamotoinstitute.org/library/bitcoin/>
- 8 Giuseppe De Ruvo, "Algorithmic Objectivity as Ideology: Toward a Critical Ethics of Digital Capitalism," *Topoi* 44 (March 2025): 175–186, <https://doi.org/10.1007/s11245-024-10117-9>
- 9 Jon Henley. "Albania puts AI-created 'minister' in charge of public procurement." *The Guardian*, September 11, 2025, <https://www.theguardian.com/world/2025/sep/11/albania-diella-ai-minister-public-procurement>
- 10 Jennifer L. Skeem and Christopher T. Lowenkamp, "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact," *Criminology* 54, no. 4 (2016): 680–712, <https://doi.org/10.1111/1745-9125.12123>.
- 11 "Predpol," Predictive Policing, Upturn, <https://teamupturn.gitbooks.io/predictive-policing/content/systems/predpol.html>.
- 12 Wikipedia, s.v. "COMPAS (Software)," [https://en.wikipedia.org/wiki/COMPAS_\(software\)](https://en.wikipedia.org/wiki/COMPAS_(software))
- 13 Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased against Blacks," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- 14 Wikipedia, s.v. "Manifest Destiny," https://en.wikipedia.org/wiki/Manifest_destiny

- 15 Marc Andreessen, “Why AI Will Save the World,” Andreessen Horowitz, June 6, 2023, <https://a16z.com/ai-will-save-the-world/>
- 16 National Security Commission on Artificial Intelligence, Final Report, March 1, 2021, 752 pp., <https://digital.library.unt.edu/ark:/67531/metadc1851188/>
- 17 Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🐦” in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ‘21), March 3-10, 2021 (virtual), 610-623, <https://doi.org/10.1145/3442188.3445922>
- 18 Steve Mollman, “OpenAI CEO Sam Altman says A.I. tools are ‘very good at doing tasks’ but terrible at doing whole jobs,” Fortune, <https://fortune.com/2023/06/08/openai-ceo-sam-altman-ai-good-at-tasks-bad-at-jobs-for-now/>
- 19 Melanie Mitchell, “Why AI is Harder Than We Think,” last modified April 28, 2021, arXiv:2104.12871, <https://arxiv.org/abs/2104.12871>
- 20 Mark Zuckerberg, “Building AGI and Open Sourcing the Journey,” Meta, January 18, 2024, <https://www.meta.com/superintelligence/>
- 21 Om Malik, “Decoding Zuck’s Superintelligence Memo,” Om Malik, July 30, 2025, <https://om.co/2025/07/30/decoding-zucks-superintelligence-memo/>
- 22 Axios, “Zuckerberg, Chan Launch \$1B Biohub to Build AI to Cure All Diseases,” November 6, 2025, <https://www.axios.com/2025/11/06/zuckerberg-chan-biohub-ai-disease>
- 23 Brian Merchant, “AI Generated Business: The Rise of AGI and the Rush to Find a Working Revenue Model,” AI Now Institute, December 2024, <https://ainowinstitute.org/publications/ai-generated-business>
- 24 Eric Schmidt and Selina Xu, “Silicon Valley Is Drifting Out of Touch With the Rest of America”, The New York Times, August 19, 2025, <https://www.nytimes.com/2025/08/19/opinion/artificial-general-intelligence-superintelligence.html>
- 25 Michael Froman, “China, the United States, and the AI Race,” Council on Foreign Relations, October 24, 2025, <https://www.cfr.org/article/china-united-states-and-ai-race>
- 26 Alex Hanna and Emily M. Bender, “AI Causes Real Harm. Let’s Focus on That over the End-of-Humanity Hype,” Scientific American, May 31, 2023, <https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/>

- 27 Elon Musk, remarks at MIT Aeronautics and Astronautics Centennial Symposium, October 24, 2014, Kresge Auditorium, Massachusetts Institute of Technology, video and summary, <https://spacepolicyonline.com/meeting-summaries/elon-musk-interviewed-at-mit-centennial-celebration-october-2014/>
- 28 Jim Collins, “Productive Paranoia,” Jim Collins, 2025, <https://www.jimcollins.com/concepts/productive-paranoia.html>
- 29 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), <https://archive.org/details/zuboff-shoshana.-the-age-of-surveillance-capitalism.-2019>
- 30 Elon Musk and Neuralink, “An Integrated Brain-Machine Interface Platform With Thousands of Channels,” *Journal of Medical Internet Research* 21, no. 10 (October 31, 2019): e16194, <https://www.jmir.org/2019/10/e16194/>
- 31 Sam Altman, “Planning for AGI and beyond,” OpenAI, February 24, 2023, <https://openai.com/index/planning-for-agi-and-beyond/>
- 32 OpenAI, “About,” OpenAI, <https://openai.com/about/>
- 33 OpenAI, “Our structure,” OpenAI, October 28, 2025, <https://openai.com/our-structure/>
- 34 Russell Brandom, “OpenAI completes its for-profit recapitalization,” *TechCrunch*, October 28, 2025, <https://techcrunch.com/2025/10/28/openai-completes-its-for-profit-recapitalization/>
- 35 OpenAI, “Introducing ChatGPT,” OpenAI, November 30, 2022, <https://openai.com/index/chatgpt/>
- 36 OpenAI, “Preparedness Framework,” OpenAI, December 18, 2023, <https://openai.com/safety/preparedness>.
- 37 Michael Feffer, Anusha Sinha, Wesley Hanwen Deng, Zachary C. Lipton & Hoda Heidari, *Red-Teaming for Generative AI: Silver Bullet or Security Theater?*, arXiv, Jan 29 2024, <https://arxiv.org/abs/2401.15897>
- 38 Karen Hao, *Empire of AI* (Allen Lane, 2025), <https://www.penguin.co.uk/books/460331/empire-of-ai-by-hao-karen/9780241678923>
- 39 The Nobel Prize in Chemistry 2024, NobelPrize.org, <https://www.nobelprize.org/prizes/chemistry/2024/summary/>
- 40 Daniela Amodei (Anthropic), “Constitutional AI,” Stanford eCorner (video transcript), <https://ecorner.stanford.edu/clips/constitutional-ai/>
- 41 Yuntao Bai et al., *Constitutional AI: Harmlessness from AI Feedback*, Anthropic, December 15, 2022, arXiv preprint, <https://arxiv.org/abs/2212.08073>

- 42 Dario Amodei, “Core Views on AI Safety: When, Why, What, and How,” Anthropic, April 2023, <https://www.anthropic.com/news/core-views-on-ai-safety>.
- 43 Yoshua Bengio, “AI for Social Good” (blog series/resource), YoshuaBengio.org, 2021. <https://yoshuabengio.org/category/ai-for-social-good/>
- 44 Future of Life Institute, “Pause Giant AI Experiments: An Open Letter,” March 22, 2023, Future of Life Institute, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- 45 U.S. House Committee on Oversight and Accountability, “Oversight of A.I.: Rules for Artificial Intelligence” (hearing before the Subcommittee on Privacy, Technology, and the Law, Senate Judiciary Committee, May 16, 2023), Gibson Dunn public policy alert (June 6, 2023), <https://www.gibsondunn.com/wp-content/uploads/2023/06/oversight-of-ai-rules-for-artificial-intelligence-and-artificial-intelligence-in-government-hearings.pdf>
- 46 Brendan Bordelon, “The law firm acting as OpenAI’s sherpa in Washington,” POLITICO Pro, September 12, 2023, <https://subscriber.politicopro.com/article/2023/09/the-law-firm-acting-as-openais-sherpa-in-washington-00115205>
- 47 Irish Council for Civil Liberties, “Scientists call on the President of the European Commission to retract AI hype statement,” *Irish Council for Civil Liberties*, November 10, 2025, <https://www.iccl.ie/press-release/scientists-call-on-the-president-of-the-european-commission-to-retract-ai-hype-statement/>
- 48 Ursula von der Leyen, “Speech by President von der Leyen at the Annual EU Budget Conference 2025,” European Commission, May 19, 2025, https://ec.europa.eu/commission/presscorner/detail/en/speech_25_1284
- 49 Will Knight, “Many Top AI Researchers Get Financial Backing From Big Tech,” WIRED, October 4 2020, <https://www.wired.com/story/top-ai-researchers-financial-backing-big-tech/>
- 50 Marc Andreessen, “Why AI Will Save the World,” Andreessen Horowitz, June 6, 2023, <https://a16z.com/ai-will-save-the-world/>
- 51 Chris Anderson, TED Talks: The Official TED Guide to Public Speaking (Boston: Houghton Mifflin Harcourt, 2016), https://archive.org/details/tedtalksofficial0000ande_f3k1
- 52 Gary Shapiro, *The Comeback: How Innovation Will Restore the American Dream* (New York: Beaufort Books, January 6, 2011), https://archive.org/details/comeback0000shap_members.iedconline.org+15
- 53 Clayton M. Christensen, *The Innovator’s Dilemma: When New Technologies Cause Great Firms to Fail* (Boston: Harvard

- Business School Press, 1997), <https://archive.org/details/innovatorsdilem000chri>
- 54 Brad Stone, *The Upstarts: How Uber, Airbnb, and the Killer Companies of the New Silicon Valley Are Changing the World* (New York: Little, Brown and Company, 2017), <https://archive.org/details/upstartshowubera0000ston>
- 55 Leigh Gallagher, *The Airbnb Story: How Three Ordinary Guys Disrupted an Industry, Made Billions, and Created Plenty of Controversy* (Boston: Houghton Mifflin Harcourt, 2017), <https://archive.org/details/airbnbstoryhowth0000gall>
- 56 Steven Levy, *In the Plex: How Google Thinks, Works, and Shapes Our Lives* (New York: Simon & Schuster, 2011), https://www.mondothèque.be/wiki/images/2/2b/Levy_Steve_In_The_Plex.pdf
- 57 Roger McNamee, *Zucked: Waking Up to the Facebook Catastrophe* (New York: Penguin Press, 2019), <https://archive.org/details/zucked-waking-up-to-the-facebook-catastrophe-2019-roger-mc-namee>
- 58 Eric Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (New York: Basic Books, 2019), https://archive.org/details/theaspen-Deep_Medicine_-_How_Artificial_Intelligence_Can_Make_Health_Care_Human_Again
- 59 Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishers, 2016), https://archive.org/details/weapons-of-math-destruction_202411
- 60 Alec MacGillis, *Fulfillment: Winning and Losing in One-Click America* (New York: Farrar, Straus and Giroux, 2021), <https://archive.org/details/fulfillmentwinni0000macg> (archive.org)
- 61 Sarah Brayne, *Predict and Surveil: Data, Discretion, and the Future of Policing* (New York: Oxford University Press, 2021), <https://academic.oup.com/book/33466> (Oxford Academic)
- 62 Ben Williamson, *Big Data in Education: The Digital Future of Learning, Policy and Practice* (London: SAGE Publications, 2017), <https://us.sagepub.com/en-us/nam/big-data-in-education/book249086>

Chapter 6

- 1 Wikipedia, s.v. “Scapegoat,” <https://en.wikipedia.org/wiki/Scapegoat>
- 2 David H. Autor, “Why Are There Still So Many Jobs? The History and Future of Workplace Automation,” *Journal of Economic Perspectives* 29, no. 3 (2015): 3–30, <https://www.aeaweb.org/articles?id=10.1257/jep.29.3.3>

- 3 Robert D. Putnam, *Bowling Alone: The Collapse and Revival of American Community* (New York: Simon & Schuster, 2000), <https://www.simonandschuster.com/books/Bowling-Alone-Revised-and-Updated/Robert-D-Putnam/9780743219037>
- 4 Martin Gilens and Benjamin I. Page, "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens," *Perspectives on Politics* 12, no. 3 (2014): 564–581, <https://www.cambridge.org/core/journals/perspectives-on-politics/article/testing-theories-of-american-politics-elites-interest-groups-and-average-citizens/62327F513959D0A304D4893B382B992B>
- 5 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014), <https://global.oup.com/academic/product/superintelligence-9780198739838>
- 6 Stanley Cohen, *Folk Devils and Moral Panics* (London: MacGibbon and Kee, 1972), <https://www.routledge.com/Folk-Devils-and-Moral-Panics/Cohen/p/book/9780415610162>
- 7 Carl Benedikt Frey and Michael A. Osborne, "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114 (2017): 254–280, <https://doi.org/10.1016/j.techfore.2016.08.019>
- 8 Jack Kelly, "Goldman Sachs Predicts 300 Million Jobs Will Be Lost Or Degraded By Artificial Intelligence," *Forbes*, March 31, 2023, <https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/>
- 9 Emily M. Bender and Alex Hanna, "The AI Con: How to Fight Big Tech's Hype and Create the Future We Want," *The Con*, <https://thecon.ai/>
- 10 Joel Mokyr, Chris Vickers, and Nicolas L. Ziebarth, "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?" *Journal of Economic Perspectives* 29, no. 3 (2015): 31–50, <https://www.aeaweb.org/articles?id=10.1257/jep.29.3.31>
- 11 David H. Autor, Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics* 118, no. 4 (2003): 1279–1333, <https://doi.org/10.1162/003355303322552801>
- 12 Simon Sharwood, "AWS CEO says using AI to replace junior staff is 'Dumbest thing I've ever heard'," *The Register*, August 21, 2025, https://www.theregister.com/2025/08/21/aws_ceo_entry_level_jobs_opinion/

- 13 Daron Acemoglu and Pascual Restrepo, "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy* 128, no. 6 (2020): 2188–2244, <https://doi.org/10.1086/705716>
- 14 Chris Mills Rodrigo, "How Major Labor Unions are Positioning on AI," Tech Policy Press, October 28, 2025, <https://techpolicy.press/how-major-labor-unions-are-positioning-on-ai>
- 15 Kevin Roose, "An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy," *New York Times*, September 2, 2022, <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>
- 16 Walter Benjamin, "The Work of Art in the Age of Mechanical Reproduction," in *Illuminations*, ed. Hannah Arendt (New York: Schocken Books, 1968), <https://web.mit.edu/allanmc/www/benjamin.pdf>
- 17 Susan Sontag, *On Photography* (New York: Farrar, Straus and Giroux, 1977), <https://www.amazon.com/Photography-Susan-Sontag/dp/0312420099>
- 18 Glyn Moody, "European Publishers Council Stays True to the Tired Old Trope about 'Copyright Theft'," *Walled Culture*, May 28, 2025, <https://walledculture.org/european-publishers-council-stays-true-to-the-tired-old-trope-about-copyright-theft/>
- 19 Carys J. Craig, *The AI-Copyright Trap* (Osgoode Legal Studies Research Paper No. 4905118, July 15, 2024), <https://ssrn.com/abstract=4905118>
- 20 Carys J. Craig, *The AI-Copyright Trap* (Osgoode Legal Studies Research Paper No. 4905118, July 15, 2024), <https://ssrn.com/abstract=4905118>
- 21 Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton, NJ: Princeton University Press, 2024), <https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil>
- 22 Pamela Samuelson, "Does Using In-Copyright Works as Training Data Infringe?" *Wolters Kluwer Copyright Blog*, July 18, 2025, <https://legalblogs.wolterskluwer.com/copyright-blog/does-using-in-copyright-works-as-training-data-infringe/>
- 23 Mark A. Lemley and Bryan Casey, *Fair Learning* (Texas Law Review, forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3528447
- 24 Nancy S. Baym, *Playing to the Crowd: Musicians, Audiences, and the Intimate Work of Connection* (New York: NYU Press, 2018), <https://nyupress.org/9781479821587/playing-to-the-crowd/>

- 25 Andres Guadamuz, "How AI is breaking traditional remuneration models," TechnoLlama, July 4, 2025, <https://www.technollama.co.uk/how-ai-is-breaking-traditional-remuneration-models>
- 26 Afzana Anwer, "Economics of Streaming & the Rise of the Music Artists' Rights and Compensation," IP Business Academy, February 24, 2025, <https://ipbusinessacademy.org/economics-of-streaming-the-rise-of-the-music-artists-rights-and-compensation>
- 27 Pamela Samuelson, "Does Using In-Copyright Works as Training Data Infringe?" Wolters Kluwer Copyright Blog, July 18, 2025, <https://legalblogs.wolterskluwer.com/copyright-blog/does-using-in-copyright-works-as-training-data-infringe/>
- 28 Andres Guadamuz, "How AI is breaking traditional remuneration models," TechnoLlama, July 4, 2025, <https://www.technollama.co.uk/how-ai-is-breaking-traditional-remuneration-models>
- 29 Caroline De Cock, "Copyright, Creativity, and Generative AI: The Research Sector's Missing Seat at the Policy Table," June 11, 2025, information labs, <https://informationlabs.org/copyright-creativity-and-generative-ai-the-research-sectors-missing-seat-at-the-policy-table/>
- 30 Wikipedia, s.v. "Dead Internet theory," https://en.wikipedia.org/wiki/Dead_Internet_theory
- 31 Isobel Owusu-Sarpong, "What Is Dead Internet Theory and Could It Change the Web?," Tech.co, May 22, 2024, <https://tech.co/news/what-is-dead-internet-theory>
- 32 Jake Renzella and Vlada Rozova, "'The Dead Internet Theory' Makes Eerie Claims About an AI-Run Web. The Truth Is More Sinister," The Conversation, April 8, 2025, <https://theconversation.com/the-dead-internet-theory-makes-eerie-claims-about-an-ai-run-web-the-truth-is-more-sinister-229609>
- 33 Dani Di Placido, "Facebook's Surreal 'Shrimp Jesus' Trend, Explained," Forbes, April 28, 2024, <https://www.forbes.com/sites/danidiplacido/2024/04/28/facebooks-surreal-shrimp-jesus-trend-explained/>
- 34 Dana Srebrnik, "The Dead Internet Theory and the Silent Surge of Bots," Wix Blog, September 18, 2024, <https://www.wix.com/blog/dead-internet-theory>
- 35 Szymon Kubala, "AI and the Dead Internet Theory: Are We Heading in That Direction?," Webmakers Expert, August 20, 2024, <https://webmakers.expert/en/blog/ai-and-the-dead-internet-theory>
- 36 Julian Horsey, "Are You Talking to Bots? The Alarming Truth About the Dead Internet Theory," Geeky Gadgets, July, 1 2025 <https://www.geeky-gadgets.com/dead-internet-theory-explained/>

- 37 Stuart Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (New York: Viking, 2019), <https://www.penguinrandomhouse.com/books/566677/human-compatible-by-stuart-russell/>
- 38 Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishers, 2016), <https://dl.acm.org/doi/10.5555/3002861>
- 39 Eliezer Yudkowsky, “Artificial Intelligence as a Positive and Negative Factor in Global Risk,” in *Global Catastrophic Risks*, ed. Nick Bostrom and Milan M. Ćirković (Oxford: Oxford University Press, 2008), <https://global.oup.com/academic/product/global-catastrophic-risks-9780199606504>
- 40 Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (New York: Hachette Books, 2020), <https://www.hachettebookgroup.com/titles/toby-ord/the-precipice/9780316484893/?lens=grand-central-publishing>
- 41 Future of Life Institute, “Pause Giant AI Experiments: An Open Letter,” March 22, 2023, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- 42 Nick Bostrom, “The Alignment Problem,” in *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014), 127–145, <https://global.oup.com/academic/product/superintelligence-9780198739838>

Chapter 7

- 1 Claudio Novelli, Luciano Floridi, Giovanni Sartor, AI as Legal Persons: Past, Patterns, and Prospects, <https://a-mcc.eu/wp-content/uploads/2024/11/ssrn-5032265.pdf>
- 2 Alex Hanna and Emily M. Bender, AI Causes Real Harm. Let’s Focus on That over the End-of-Humanity Hype, *Scientific American*, <https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/>
- 3 European Parliament Committee on Legal Affairs (JURI), Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (PE 582.443v01-00, JURI-PR 582443), https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf
- 4 Alex Hern, Give robots ‘personhood’ status, EU committee argues, *The Guardian*, <https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues>
- 5 European Parliament Press Release, Robots: Legal Affairs Committee calls for EU-wide rules, <https://www.europarl.europa.eu>

- eu/news/en/press-room/20170110IPR57613/robots-legal-affairs-committee-calls-for-eu-wide-rules
- 6 Jeffrey Dastin, Insight - Amazon scraps secret AI recruiting tool that showed bias against women, Reuters, <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>
 - 7 Charlotte Lytton, AI hiring tools may be filtering out the best job applicants, BBC, <https://www.bbc.com/worklife/article/20240214-ai-recruiting-hiring-software-bias-discrimination>
 - 8 Edmund L. Andrews, How Flawed Data Aggravates Inequality in Credit, Stanford, <https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>
 - 9 Lee McIntyre, *Post-Truth* (Cambridge, MA: MIT Press, 2018), <https://mitpress.mit.edu/9780262535045/post-truth/>
 - 10 Julien Kiese Bahangulu and Louis Owusu-Berko, “Algorithmic Bias, Data Ethics, and Governance: Ensuring Fairness, Transparency and Compliance in AI-Powered Business Analytics Applications,” *World Journal of Advanced Research and Reviews* 25, no. 2 (2025): 1746–63, <https://doi.org/10.30574/wjarr.2025.25.2.0571>
 - 11 Matthew Chiu, “Path Dependency in Policymaking: A Double-Edged Sword,” Protopia Group, April 19, 2024, <https://www.protopiagroup.org/analysis/path-dependency-in-policymaking-a-double-edged-sword>
 - 12 Jacob Robbins, “AI startups grabbed a third of global VC dollars in 2024,” Pitchbook, <https://pitchbook.com/news/articles/ai-startups-grabbed-a-third-of-global-vc-dollars-in-2024>
 - 13 Sarah Perez, “Character.AI, the a16z-backed chatbot startup, tops 1.7M installs in first week,” TechCrunch, <https://techcrunch.com/2023/05/31/character-ai-the-a16z-backed-chatbot-startup-tops-1-7m-installs-in-first-week/>
 - 14 Kenrick Kai, “Google hires top talent from startup Character.AI, signs licensing deal,” Reuters, <https://www.reuters.com/technology/artificial-intelligence/google-hires-characterai-cofounders-licenses-its-models-information-reports-2024-08-02/>
 - 15 Amazon staff, “Amazon concludes \$4 billion investment in Anthropic,” <https://www.aboutamazon.com/news/company-news/amazon-anthropic-ai-investment>
 - 16 Naomi Nix, Cat Zakrzewski and Gerrit De Vynck, “Silicon Valley is pricing academics out of AI research,” The Washington Post, <https://www.washingtonpost.com/technology/2024/03/10/big-tech-companies-ai-research/>

- 17 Naomi Nix, Cat Zakrzewski, and Gerrit De Vynck, "Silicon Valley Is Pricing Academics Out of AI Research," *The Washington Post*, March 10, 2024, <https://www.washingtonpost.com/technology/2024/03/10/big-tech-companies-ai-research/>
- 18 Nicklas Berild Lundblad, "Unpredictable Patterns #115: Questioning our assumptions," *Unpredictable Patterns* (blog), April 18, 2025, <https://unpredictablepatterns.substack.com/p/unpredictable-patterns-115-questioning>
- 19 Fabrizio Gilardi and Fabio Wasserfallen, "The Politics of Policy Diffusion," *European Journal of Political Research* (2019), <https://doi.org/10.1111/1475-6765.12326>
- 20 Wikipedia, s.v. "Nirvana fallacy", https://en.wikipedia.org/wiki/Nirvana_fallacy
- 21 Audrey Watters, "The History of the Future of Calculators in Schools," *Hack Education*, March 12, 2015, <https://hackeducation.com/2015/03/12/calculators>
- 22 Elana Klein, "Tech in the Classroom: A History of Hype and Hysteria," *Wired*, August 18, 2025, <https://www.wired.com/story/photo-essay-school-tech-hysteria/>
- 23 Richard A. Clarke and Robert K. Knake, *Cyber War: The Next Threat to National Security and What to Do About It* (New York: Ecco, 2010), <https://www.harpercollins.com/products/cyber-war-richard-a-clark robertknake>
- 24 Leon E. Panetta, "Defending the Nation from Cyber Attack," *Business Executives for National Security*, October 11, 2012, <https://nsarchive.gwu.edu/sites/default/files/documents/2700165/Document-78.pdf>
- 25 Committee on the Judiciary and Select Subcommittee on the Weaponization of the Federal Government, *The Weaponization of CISA: How a "Cybersecurity" Agency Colluded with Big Tech and "Disinformation" Partners to Censor Americans: Interim Staff Report of the Committee on the Judiciary and the Select Subcommittee on the Weaponization of the Federal Government* (U.S. House of Representatives, June 26, 2023), <https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/cisa-staff-report6-26-23.pdf>
- 26 Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (New York: W. W. Norton & Company, 2015), <https://www.schneier.com/books/data-and-goliath/>
- 27 P.W. Singer and Allan Friedman, *Cybersecurity and Cyberwar: What Everyone Needs to Know* (Oxford: Oxford University Press,

- 2014), <https://global.oup.com/academic/product/cybersecurity-and-cyberwar-9780199918096>
- 28 National Academy of Sciences, *At the Nexus of Cybersecurity and Public Policy: Some Basic Concepts and Issues* (Washington, DC: National Academies Press, 2014), <https://nap.nationalacademies.org/catalog/18749/at-the-nexus-of-cybersecurity-and-public-policy-some-basic>
 - 29 Langdon Winner, "Do Artifacts Have Politics?" *Daedalus* 109, no. 1 (1980): 121–136, <https://www.jstor.org/stable/20024652>
 - 30 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), <https://www.hup.harvard.edu/books/9780674970847>
 - 31 Daniel Boffey, "EU eyes temporary ban on facial recognition in public places," *The Guardian*, January 17, 2020, <https://www.theguardian.com/technology/2020/jan/17/eu-eyes-temporary-ban-on-facial-recognition-in-public-places>
 - 32 Thomas Macaulay, "Oh great, the EU has ditched its facial recognition ban," *The Next Web*, February 12, 2020, <https://thenextweb.com/news/oh-great-the-eu-has-ditched-its-facial-recognition-ban>
 - 33 European Parliament, "Resolution on Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters," October 6, 2021, https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html
 - 34 Future of Life Institute, "Pause Giant AI Experiments: An Open Letter," March 22 2023, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
 - 35 Geoffrey Gertz and Justin Sherman, "Beyond Bans: Expanding the Policy Options for Tech-Security Threats," *Lawfare*, June 25, 2025, <https://www.lawfaremedia.org/article/beyond-bans--expanding-the-policy-options-for-tech-security-threats>
 - 36 Hal Abelson et al., "Bugs in Our Pockets: The Risks of Client-Side Scanning," October 15, 2021, arXiv, <https://doi.org/10.48550/arXiv.2110.07450>
 - 37 Karen Levy, "The Contexts of Control: Information, Power, and Truck-Driving Work," *Information Society* 31, no. 2 (2015): 160–174, <https://doi.org/10.1080/01972243.2015.998105>
 - 38 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven: Yale University Press, 2018), <https://doi.org/10.12987/9780300235029>

- 39 Solon Barocas, Moritz Hardt, and Arvind Narayanan, *Fairness and Machine Learning: Limitations and Opportunities* (Cambridge, MA: MIT Press, 2023), <https://fairmlbook.org/>
- 40 Jillian C. York, *Silicon Values: The Future of Free Speech Under Surveillance Capitalism* (London: Verso Books, 2021), <https://www.versobooks.com/products/882-silicon-values>
- 41 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018), <https://us.macmillan.com/books/9781250074317/automatinginequality>
- 42 Pauline T. Kim, "Data-Driven Discrimination at Work," *William & Mary Law Review* 58, no. 3 (2017): 857–936, <https://scholarship.law.wm.edu/wmlr/vol58/iss3/4/>
- 43 danah boyd, "The Politics of 'Real Names'," *Communications of the ACM* 55, no. 8 (2012): 29–31, <https://doi.org/10.1145/2240236.2240247>
- 44 Svea Windwehr, Jillian C. York, and Christoph Schmon, "Just Banning Minors From Social Media Is Not Protecting Them," *Electronic Frontier Foundation*, July 28, 2025, <https://www.eff.org/deeplinks/2025/07/just-banning-minors-social-media-not-protecting-them>
- 45 Sonia Livingstone, "iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy," *Journal of Children and Media* 12, no. 1 (2018): 118–123, <https://doi.org/10.1080/17482798.2017.1417091>
- 46 Electronic Frontier Foundation, "Age Verification Mandates Would Undermine Anonymous Speech and Require Extensive Data Collection," March 2023, <https://www.eff.org/deeplinks/2023/03/age-verification-mandates-would-undermine-anonymous-speech-and-require-extensive>
- 47 Jeff Kosseff, *The Twenty-Six Words That Created the Internet* (Ithaca: Cornell University Press, 2019), <https://www.cornellpress.cornell.edu/book/9781501714412/the-twenty-six-words-that-created-the-internet/>
- 48 Rebecca MacKinnon, *Consent of the Networked: The Worldwide Struggle for Internet Freedom* (New York: Basic Books, 2012), <https://consentofthenetworked.com/>
- 49 Sarah T. Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven: Yale University Press, 2019), <https://www.jstor.org/stable/j.ctvhrcz0v>
- 50 Joanna Bryson, "AI Regulation and the Limits of Transparency," *AXA Research Fund Insight* (July 12, 2021), <https://www.axa.com/en/insights/ai-regulation-and-the-limits-of-transparency>.

- 51 “Predpol,” Predictive Policing, Upturn, <https://teamupturn.gitbooks.io/predictive-policing/content/systems/predpol.html>.
- 52 Wikipedia, s.v. “COMPAS (Software),” [https://en.wikipedia.org/wiki/COMPAS_\(software\)](https://en.wikipedia.org/wiki/COMPAS_(software))

Chapter 8

- 1 Nick Lichtenberg, “Is AI a Bubble? Market Crash Sees \$1 Trillion Wiped From S&P 500 as OpenAI’s GPT-5 Release Flops and Sam Altman’s ‘Bubble’ Comments Bite,” *Fortune*, August 24, 2025, <https://fortune.com/2025/08/24/is-ai-a-bubble-market-crash-gary-marcus-openai-gpt5/>
- 2 Robert Brauneis, “Copyright Law and New Technologies: A Long and Complex Relationship,” *Unfolding the Past* (blog), Library of Congress, May 18, 2017, <https://blogs.loc.gov/copyright/2017/05/copyright-law-and-new-technologies-a-long-and-complex-relationship/>
- 3 Wikipedia, s.v. “White-Smith Music Publishing Co. v. Apollo Co.,” https://en.wikipedia.org/wiki/White-Smith_Music_Publishing_Co._v._Apollo_Co.
- 4 The Copyright Act of 1909 (as amended and codified), Title 17 U.S.C., https://ipmall.law.unh.edu/sites/default/files/hosted_resources/lipa/copyrights/CopyrightAct1909amendedcodified_.pdf
- 5 Zvi S. Rosen, “The Unoriginal Sin of Technology-Neutral Copyright,” *Minnesota Law Review* 100 (2016): 1495–1558, <https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1206&context=mlr>
- 6 Imagining the Internet | Elon University. “1830s – 1860s: Telegraph,” n.d. <https://www.elon.edu/u/imagining/time-capsule/150-years/back-1830-1860/>
- 7 Tim Wu, “The Master Switch: The Rise and Fall of Information Empires” (working paper, Columbia Law School, 2007), https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=2462&context=faculty_scholarship
- 8 United States Congress, “Federal Criminal Code, Title 18, Part I, Chapter 63 & Chapter 113”, https://www.oas.org/juridico/spanish/cyber/us_cyb_law_fed_crim_code.pdf
- 9 LII / Legal Information Institute. “Wire Fraud,” n.d. https://www.law.cornell.edu/wex/wire_fraud
- 10 Investopedia. “Understanding Wire Fraud: Definition, Laws, and Key Examples,” August 27, 2025. <https://www.investopedia.com/terms/w/wirefraud.asp>
- 11 Valerie C. Brannon and Eric N. Holmes, Section 230: An Overview, CRS Report R46751 (Washington, D.C.: Congressional

- Research Service, January 4 2024), <https://www.congress.gov/crs-product/R46751>
- 12 Electronic Frontier Foundation. "Section 230," n.d. <https://www.eff.org/issues/cda230>
 - 13 Electronic Frontier Foundation, "Section 230," <https://www.eff.org/issues/cda230>; see also "What you should know about Section 230, the rule that shaped today's internet," PBS News, December 2021, <https://www.pbs.org/newshour/politics/what-you-should-know-about-section-230-the-rule-that-shaped-todays-internet>
 - 14 European Parliament and Council, Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, Official Journal L 178 (17 July 2000): 1-16, <https://eur-lex.europa.eu/eli/dir/2000/31/oj/eng>
 - 15 Wikipedia, s.v. "The Electronic Commerce Directive (2000/31/EC)," https://en.wikipedia.org/wiki/Electronic_Commerce_Directive_2000
 - 16 "House of Lords - Regulating in a Digital World - Select Committee on Communications," n.d. <https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/29908.htm>
 - 17 CEPS, "Limited liability for the net? The Future of Europe's E-Commerce Directive," 2020, <https://www.ceps.eu/ceps-publications/limited-liability-net-future-europes-e-commerce-directive/>
 - 18 Matthijs Maas, "AI is like...: A Literature Review of AI Metaphors and Why They Matter for Policy," Institute for Law & AI, AI Foundations Report 2, October 2023, <https://law-ai.org/ai-policy-metaphors/>
 - 19 Britta C. Brugman, Christian Burgers, and Barbara Vis. "Metaphorical Framing in Political Discourse through Words vs. Concepts: A Meta-Analysis." *Language and Cognition* 11, no. 1 (March 1, 2019): 41–65. <https://doi.org/10.1017/langcog.2019.5>
 - 20 Centre for Trustworthy Technology, "Metaphors Matter: How Language Influences Technology and Society," 2024, <https://c4tt.org/metaphors-matter-how-language-influences-technology-and-society/>
 - 21 Keith Rasenberger, "Why Metaphors Matter in AI Law and Policy," *The Regulatory Review*, May 2, 2024, <https://www.theregview.org/2024/05/02/rasenberger-why-metaphors-matter-in-ai-law-and-policy/>
 - 22 Andres Guadamuz, "Artists file class-action lawsuit against Stability AI, DeviantArt, and Midjourney," *TechnoLlama*,

- January 15, 2023, <https://www.technollama.co.uk/artists-file-class-action-lawsuit-against-stability-ai-deviantart-and-midjourney>
- 23 Wikipedia, s.v. “Semantic gap,” https://en.wikipedia.org/wiki/Semantic_gap
 - 24 Jeffrey A. Frankel, “Over-optimism in Forecasts by Official Budget Agencies and Its Implications,” *Oxford Review of Economic Policy* 27, no. 4 (Winter 2011): 536-562, <https://www.hks.harvard.edu/publications/over-optimism-forecasts-official-budget-agencies-and-its-implications>
 - 25 Saul Hudson et al., “Policy failure and the policy-implementation gap: can policy support programs help?” *Policy Studies* 40, no. 1 (2019): 129-148, <https://www.tandfonline.com/doi/full/10.1080/25741292.2018.1540378>
 - 26 Landry Signé and Colette van der Ven, “Turning policy into action: Overcoming policy failures and bridging implementation gaps,” *Brookings*, February 28, 2024, <https://www.brookings.edu/articles/turning-policy-into-action-in-africa/>
 - 27 Thomas H. Beach, Yacine Y. Rezgui, Haijiang Li, and T. Kasim. “A Rule-Based Semantic Approach for Automated Regulatory Compliance in the Construction Sector.” *Expert Systems With Applications* 42, no. 12 (March 6, 2015): 5219–31. <https://doi.org/10.1016/j.eswa.2015.02.029>
 - 28 Suzy E. Park, *Technological Convergence: Regulatory, Digital Privacy, and Data Security Issues*, CRS Report R45746 (Washington, D.C.: Congressional Research Service, May 30, 2019), https://www.everycrsreport.com/files/20190530_R45746_461fe45bf0ce4a3f25e8d1a507c308fee4d5bfd4.pdf
 - 29 “Closing the ‘RegLag’ Between Prospective Regulated Activity and Regulation,” *CLS Blue Sky Blog*, October 16, 2020, <https://clsbluesky.law.columbia.edu/2020/10/16/closing-the-reglag-between-prospective-regulated-activity-and-regulation/>
 - 30 OECD, *Case Studies on the Regulatory Challenges Raised by Innovation and the Regulatory Responses* (Paris: OECD Publishing, 2023), https://www.oecd.org/en/publications/case-studies-on-the-regulatory-challenges-raised-by-innovation-and-the-regulatory-responses_8fa190b5-en.html
 - 31 Carlos F. Ceballos Ferroglio, Gustavo Ferro, and Ángel Enrique Neder, “Regulatory Lag, Efficiency, and Performance: Lessons from a Case Study,” *Competition and Regulation in Network Industries* 25, no. 1 (2024): 43-66, <https://doi.org/10.1177/17835917241251811>

- 32 Emilio Kyprianou, “The Red Flag Act,” The Open University Law School Blog, October 2 2023, <https://law-school.open.ac.uk/blog/red-flag-act>
- 33 Winston J. Maxwell and Marc Bourreau, “Technology Neutrality in Internet, Telecoms and Data Protection Regulation” (Computer and Telecommunications Law Review, 2014), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2529680

Chapter 9

- 1 Langdon Winner, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought* (Cambridge, MA: MIT Press, 1977), <https://mitpress.mit.edu/9780262230780/autonomous-technology/>
- 2 Helen Nissenbaum, “How Computer Systems Embody Values,” *Computer* 34, no. 3 (2001): 120–119, <https://doi.org/10.1109/2.910905>
- 3 Myra Cheng et al., “From Tools to Thieves: Measuring and Understanding Public Perceptions of AI through Crowdsourced Metaphors,” arXiv, January 31, 2025, <https://arxiv.org/html/2501.18045v1>
- 4 Madhumita Murgia, interview of Ted Chiang, “Sci-fi writer Ted Chiang: ‘The machines we have now are not conscious’,” *Financial Times*, June 3–4, 2023, <https://www.ft.com/content/c1f6d948-3dde-405f-924c-09cc0dcf8c84>
- 5 Melissa Heikkilä, “Why it’ll be hard to tell if AI ever becomes conscious,” *MIT Technology Review*, https://www.technologyreview.com/2023/10/17/1081818/why-itll-be-hard-to-tell-if-ai-ever-becomes-conscious/?utm_source=LinkedIn
- 6 Kevin Roose, “If A.I. Systems Become Conscious, Should They Have Rights?”, *The New York Times*, <https://www.nytimes.com/2025/04/24/technology/ai-welfare-anthropoc-claude.html>
- 7 Ryan Calo, “Products Liability Law as a Way to Address AI Harms,” *Brookings Institution*, October 31, 2019, <https://www.brookings.edu/articles/products-liability-law-as-a-way-to-address-ai-harms/>
- 8 Wikipedia, s.v. “MacPherson v. Buick Motor Co.,” https://en.wikipedia.org/wiki/MacPherson_v._Buick_Motor_Co.
- 9 W. Kerrel Murray, “Products Liability for Artificial Intelligence,” *Lawfare*, May 1, 2024, <https://www.lawfaremedia.org/article/products-liability-for-artificial-intelligence>
- 10 Ryan Calo, “Robotics and the Lessons of Cyberlaw,” *California Law Review* 103, no. 3 (2015): 513–563, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2402972

- 11 Michael J. M. L. W. Horton, "Artificial Intelligence & Product Liability," McCarter & English, LLP, September 26, 2023, <https://www.mccarter.com/insights/artificial-intelligence-product-liability/>
- 12 "EU AI Act: First Regulation on Artificial Intelligence," European Parliament, February 19, 2025, <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- 13 Julie E. Cohen, "The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy," *Philosophy & Technology* 31, no. 2 (2018): 213–233, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2666570
- 14 Albert Borgmann, *Holding On to Reality: Digital Media and the Path to Understanding* (Chicago: University of Chicago Press, 2006), <https://press.uchicago.edu/ucp/books/book/chicago/H/bo3640475.html>
- 15 Leonardo/ISASTwith Arizona State University. "Review of Privacy's Blueprint: The Battle to Control the Design of New Technologies," April 29, 2021. <https://leonardo.info/review/2018/12/review-of-privacys-blueprint-the-battle-to-control-the-design-of-new-technologies>
- 16 Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science* 356, no. 6334 (2017): 183–186, <https://www.science.org/doi/10.1126/science.aal4230>
- 17 Oishee Kundu, Elvira Uyarra, Raquel Ortega-Argiles, Mayra M Tirado, Tasos Kitsos, Pei-Yu Yuan, Impacts of policy-driven public procurement: a methodological review, *Science and Public Policy*, Volume 52, Issue 1, February 2025, Pages 50–64, <https://doi.org/10.1093/scipol/scae058>
- 18 Batya Friedman, Peter H. Kahn Jr., and Alan Borning, "Value Sensitive Design and Information Systems," in *The Handbook of Information and Computer Ethics*, ed. Kenneth Einar Himma and Herman T. Tavani (Hoboken, NJ: John Wiley & Sons, 2008), 69–101, https://android.eng.ankara.edu.tr/wp-content/uploads/sites/656/2017/10/10_The_Handbook_of_Information_and_Computer_Ethics.pdf
- 19 Peter-Paul Verbeek, *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (Penn State University Press, 2005), <https://www.jstor.org/stable/10.5325/j.ctv14gp4w7>
- 20 Andrew McAfee and Erik Brynjolfsson, *Machine, Platform, Crowd: Harnessing Our Digital Future* (New York: W. W. Norton & Company, 2017), <https://www.amazon.com/Machine-Platform-Crowd-Harnessing-Digital/dp/0393254291>

- 21 Deirdre K. Mulligan, Joshua A. Kroll, Nitin Kohli, and Richmond Y. Wong, "This Thing Called Fairness," *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (November 7, 2019): 1–36. <https://doi.org/10.1145/3359221>
- 22 Frank Pasquale, "The Algorithmic Self," *The Hedgehog Review* 17, no. 1 (Spring 2015). Available at: <https://hedgehogreview.com/issues/too-much-information/articles/the-algorithmic-self>
- 23 Maja van der Velden and Christina Mörtberg, "Participatory Design and Design for Values," in *Handbook of Ethics, Values, and Technological Design*, ed. Jeroen van den Hoven, Peter E. Vermaas, and Ibo van de Poel (Dordrecht: Springer, 2015), 41–66, https://doi.org/10.1007/978-94-007-6970-0_33
- 24 Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* 3, no. 1 (2016): 1–12, <https://journals.sagepub.com/doi/full/10.1177/2053951715622512>
- 25 Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford: Oxford University Press, 2016), <https://global.oup.com/academic/product/technology-and-the-virtues-9780190498511>
- 26 John Cheney-Lippold, *We Are Data: Algorithms and the Making of Our Digital Selves* (New York: NYU Press, 2017), <https://www.jstor.org/stable/j.ctt1gk0941>
- 27 Lina M. Khan, "Amazon's Antitrust Paradox," *Yale Law Journal* 126, no. 3 (2017): 710–805, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2911742
- 28 Mary L. Gray and Siddharth Suri, *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Boston: Houghton Mifflin Harcourt, 2019), <https://www.microsoft.com/en-us/research/publication/ghost-work-how-to-stop-silicon-valley-from-building-a-new-global-underclass-co-authored-with-siddharth-suri/>
- 29 Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford: Oxford University Press, 2020), <https://academic.oup.com/book/36491>
- 30 MDR Regulator, "Clinical trials of medical devices confirm that the products are designed and manufactured in accordance with legal requirements," MDR Regulator: CE Marking for Medical Devices, <https://mdrregulator.com/ce-marking/clinical-trials-of-medical-devices>
- 31 Michael D. Abràmoff, Philip T. Lavin, Michele Birch, Nilay Shah, and James C. Folk, "Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in

- Primary Care Offices,” *npj Digital Medicine* 1 (2018): 39, <https://doi.org/10.1038/s41746-018-0040-6>
- 32 Guido D. Giebel et al., “Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study,” *Journal of Medical Internet Research* 27 (2025): e63377, <https://doi.org/10.2196/63377>
- 33 British Medical Association, “Principles for Artificial Intelligence (AI) and Its Application in Healthcare” (London: British Medical Association, 2024), <https://www.bma.org.uk/media/njgfbmnn/bma-principles-for-artificial-intelligence-ai-and-its-application-in-healthcare.pdf>
- 34 Casey Ross and Ike Swetlitz, “IBM Pitched Its Watson Supercomputer as a Revolution in Cancer Care. It’s Nowhere Close,” *STAT*, September 5, 2017, <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>
- 35 Google DeepMind, “AlphaFold: a solution to a 50-year-old grand challenge in biology,” <https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/>
- 36 Luciana DeBenedetti, “Togo’s Novissi Cash Transfer: Designing and Implementing a Fully Digital Social Assistance Program during COVID-19,” *Innovations for Poverty Action*, August 4, 2021, <https://poverty-action.org/publication/togo%E2%80%99s-novissi-cash-transfer-designing-and-implementing-fully-digital-social-assistance>
- 37 Emily Aiken, Suzanne Bellue, Dean Karlan, Christopher Udry, and Joshua E. Blumenstock, “Machine Learning and Phone Data Can Improve Targeting of Humanitarian Aid,” *Nature* 603, no. 7903 (March 31, 2022): 864–70, <https://doi.org/10.1038/s41586-022-04484-9>.
- 38 International Civil Aviation Organisation (ICAO), <https://www.icao.int/Pages/default.aspx>
- 39 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), <https://www.ich.org/>
- 40 Tim Wu, “A Brief History of American Telecommunications Regulation,” *Oxford International Encyclopedia Of Legal History*, Vol. 5, p. 95, 2009 (2007), https://scholarship.law.columbia.edu/faculty_scholarship/1461/
- 41 U.S. Food & Drug Administration, “The Drug Development Process – Step 3: Clinical Research,” <https://www.fda.gov/patients/drug-development-process/step-3-clinical-research>
- 42 U.S. National Highway Traffic Safety Administration, “Laws & Regulations,” <https://www.nhtsa.gov/laws-regulations>

- 43 Government of Canada, “Algorithmic Impact Assessment tool,” Canada.ca, <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>
- 44 Treasury Board of Canada Secretariat, Algorithmic Impact Assessment Tool, Government of Canada, <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>
- 45 Serena Oduro, Emanuel Moss, and Jacob Metcalf, “Obligations to Assess: Recent Trends in AI Accountability Regulations,” *Patterns* 3, no. 11 (November 11, 2022): 100608, <https://doi.org/10.1016/j.patter.2022.100608>
- 46 U.K. Financial Conduct Authority, “Regulatory Sandbox,” <https://www.fca.org.uk/firms/innovation/regulatory-sandbox>
- 47 Wikipedia, s.v. “Artificial Intelligence Act,” https://en.wikipedia.org/wiki/Artificial_Intelligence_Act
- 48 New York State Assembly bill A6453A, Relates to the training and use of artificial intelligence frontier models, 2025-2026 Leg. Sess. (N.Y. 2025), <https://www.nysenate.gov/legislation/bills/2025/A6453/amendment/A>
- 49 California SB 53, Artificial intelligence models: large developers, 2025-2026 Reg. Sess. (Cal. 2025), https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB53
- 50 Kevin Frazier and Andrew W. Reddie, “Why Context, Not Compute, is the Key to AI Governance,” Tech Policy Press, August 5, 2025, <https://www.techpolicy.press/why-context-not-compute-is-the-key-to-ai-governance/>
- 51 U.S. Congress, Office of Technology Assessment, The OTA Legacy, Princeton University (archived), <http://www.princeton.edu/~ota/>
- 52 Bruce A. Bimber, *The Politics of Expertise in Congress: The Rise and Fall of the Office of Technology Assessment* (Albany, NY: State University of New York Press, 1996), <https://archive.org/details/politicsofexpert0000bimb>
- 53 U.S. Congressional Budget Office, “Introduction to CBO,” <https://www.cbo.gov/about/overview>
- 54 U.S. Food & Drug Administration, “Breakthrough Devices Program,” <https://www.fda.gov/medical-devices/how-study-and-market-your-device/breakthrough-devices-program>
- 55 Environmental Protection Agency, “Toxics Release Inventory (TRI) Program,” U.S. Environmental Protection Agency, last

- updated July 3, 2025, <https://www.epa.gov/toxics-release-inventory-tri-program/>
- 56 Bank for International Settlements, “CAP10 Definition of eligible capital,” in Basel Framework, https://www.bis.org/basel_framework/chapter/CAP/10.htm docs.oracle.com+15
 - 57 United Nations Environment Programme, “The Montreal Protocol on Substances that Deplete the Ozone Layer,” <https://ozone.unep.org/treaties/montreal-protocol>
 - 58 Global Partnership on Artificial Intelligence, “2024 GPAI New Delhi Declaration,” adopted July 3-4, 2024, New Delhi, para. 1-2, <https://gpai.ai/gpai-new-delhi-declaration-2024.pdf>
 - 59 European Parliamentary Research Service, Digital Sovereignty for Europe, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI\(2020\)651992_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf)
 - 60 Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (New York: Oxford University Press, 2020), 189, <https://global.oup.com/academic/product/the-brussels-effect-9780190088583>
 - 61 Caroline De Cock, “Digital Sovereignty: Europe’s Favourite (Empty) Buzzword,” Information Labs (blog), October 22, 2025, <https://informationlabs.org/digital-sovereignty-europes-favourite-empty-buzzword/>
 - 62 Ursula von der Leyen, “A Union that strives for more. My agenda for Europe - Political Guidelines for the Next Commission,” European Commission, , https://commission.europa.eu/document/download/063d44e9-04ed-4033-acf9-639ecb187e87_en?filename=political-guidelines-next-commission_en.pdf
 - 63 Neil Postman, “Technopoly: The Surrender of Culture to Technology”, Vintage Books, 1993, <https://ia601802.us.archive.org/27/items/235p-technopoly-neil-postman/235p%20technopoly-neil-postman.pdf>
 - 64 Brian Walker and David Salt, *Resilience Thinking: Sustaining Ecosystems and People in a Changing World* (Washington, DC: Island Press, 2006), <https://www.amazon.com/Resilience-Thinking-Sustaining-Ecosystems-Changing/dp/1597260932>
 - 65 Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press, 1990), <https://www.amazon.com/Governing-Commons-Evolution-Institutions-Collective/dp/0521405998>
 - 66 James F. Tracy, “Review: C. Edwin Baker, *Media Concentration and Democracy: Why Ownership Matters*. New York: Cambridge University Press, 2007. £35.00 (Hbk), £14.99 (Pbk). 256 Pp.”

- European Journal of Communication 23, no. 2 (May 30, 2008): 249–53. <https://doi.org/10.1177/02673231080230020610>
- 67 James S. Coleman, “Social Capital in the Creation of Human Capital,” *American Journal of Sociology* 94 (1988): S95–S120, <http://www.jstor.org/stable/2780243>
- 68 European Commission, “AI Act Implementation Timeline,” European AI Act Timeline, <https://artificialintelligenceact.eu/implementation-timeline/>
- 69 U.S. Food & Drug Administration, Classify Your Medical Device, FDA, <https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device>
- 70 U.S. Food & Drug Administration, Safety Management System (SMS), Federal Aviation Administration, <https://www.faa.gov/about/initiatives/sms>.
- 71 Amba Kak and Sarah West, 2. A Modern Industrial Strategy for AI?: Interrogating the US Approach - AI Now Institute, <https://ainowinstitute.org/publication/a-modern-industrial-strategy-for-aiinterrogating-the-us-approach>
- 72 Ed Morrison, “Complexity Thinking. What Is It?” Strategic Doing Institute (August 30, 2023), <https://agilestrategylab.org/complexity-thinking-what-is-it/>
- 73 Beth Noveck, “Crowdlaw: Collective Intelligence and Lawmaking,” *Analyse & Kritik* 40, no. 2 (2018): 359–380, https://www.researchgate.net/publication/328728004_Crowdlaw_Collective_Intelligence_and_Lawmaking
- 74 Guy Standing, *Work After Globalization: Building Occupational Citizenship* (Cheltenham, UK: Edward Elgar Publishing, 2009), <https://www.elgaronline.com/monobook/9781848441644.xml>
- 75 Julia Black, “The Rise, Fall and Fate of Principles Based Regulation,” LSE Law, Society and Economy Working Papers 13/2008, London School of Economics and Political Science, 2008, <https://www.lse.ac.uk/law/research/working-paper-series/2007-08/WPS2008-13-Black.pdf>
- 76 “Principles-Based Regulations: What Are the Opportunities and Trade-Offs?” Australian Energy Council, May 22, 2024, <https://www.energycouncil.com.au/analysis/principles-based-regulations-what-are-the-opportunities-and-trade-offs/>
- 77 “Uniform Commercial Code,” Uniform Law Commission, <https://www.uniformlaws.org/acts/ucc>
- 78 David M. Farrell, Jane Suiter, and Clodagh Harris, “‘Systematizing’ Constitutional Deliberation: The 2016–18 Citizens’ Assembly in Ireland,” *Irish Political*

- Studies 34, no. 1 (January 2019): 113–23, <https://doi.org/10.1080/07907184.2018.1534832>
- 79 Lars Klüver, “Consensus Conferences at the Danish Board of Technology,” in *Public Participation in Science: The Role of Consensus Conferences in Europe*, ed. Simon Joss and John Durant (London: Science Museum, 1995), 42–53, <https://www.jstor.org/stable/20059945>.
- 80 Participedia, “Participatory Consensus Conferences,” Participedia, <https://participedia.net/method/participatory-consensus-conferences>
- 81 Dr. Dominik Hierlemann and Dr. Stefan Roch, *Digital Democracy – What Europe Can Learn from Taiwan* (Gütersloh: Bertelsmann Stiftung, September 2020), https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Demokratie_und_Partizipation_in_Europa_/210211_Bericht_Digital_Democracy_-_What_Europe_can_learn_from_Taiwan___1_.pdf
- 82 Gemeente Amsterdam, *Algoritmeregister*, last modified June 4, 2025, <https://algoritmes.overheid.nl/nl/organisatie/gm0363/gemeente-amsterdam>
- 83 Municipality of Amsterdam, *Grip on algorithms: Approach and tools for a responsible use of algorithms in Amsterdam* (Amsterdam: City of Amsterdam, June 2022), https://www.amsterdam.nl/publish/pages/1053010/playbook_algorithms.pdf
- 84 City of Helsinki, “Ethical Principles and the Algorithm Register,” *Helsinki Data Strategy*, <https://digi.hel.fi/english/helsinki-city-data-strategy/helsinki-datastrategy-chapter-5/55-chapter/>
- 85 Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique (Digital Republic Act), *Journal officiel de la République française* (8 October 2016), <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte>
- 86 Regulation (EU) 2024/1689, Arts. 12–15 (“Transparency obligations”), *Official Journal of the European Union* L 1689 (12 July 2024), in force 2 February 2025, <https://artificialintelligenceact.eu/article/13/> (Arts. 12–15).
- 87 Decidim Association, “Decidim: Free Open-Source Participatory Democracy Platform,” <https://decidim.org/>
- 88 Trebor Scholz, *Platform Cooperativism: Challenging the Corporate Sharing Economy* (New York: Rosa Luxemburg Stiftung, January 2016), https://rosalux.nyc/wp-content/uploads/2020/11/RLS-NYC_platformcoop.pdf
- 89 Ellen Hodgson Brown, *The Public Bank Solution: From Austerity to Prosperity* (Baton Rouge: Third Millennium Press, 2013), <https://archive.org/details/publicbanksoluti0000brow>.

- 90 Tarmo Virki, "Finland Seeks to Teach 1 % of All Europeans Basics on AI," Reuters, December 10, 2019, <https://www.reuters.com/article/world/finland-seeks-to-teach-1-of-all-europeans-basics-on-ai-idUSKBN1YE1CT>
- 91 Steering Group of the Artificial Intelligence Programme, Finland's Age of Artificial Intelligence: Turning Finland into a Leading Country in the Application of Artificial Intelligence, Publications of the Ministry of Economic Affairs and Employment 47/2017 (Helsinki: Ministry of Economic Affairs and Employment, October 2017), https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y
- 92 Douglas Kellner and Jeff Share, "Toward Critical Media Literacy: Core Concepts, Debates, Organizations, and Policy," *Discourse: Studies in the Cultural Politics of Education* 26, no. 3 (2005): 369–386, <https://doi.org/10.1080/01596300500200169>
- 93 AlgorithmWatch gGmbH and International Trade Union Confederation, Algorithmic Transparency and Accountability in the World of Work: A Mapping Study into the Activities of Trade Unions (Berlin: AlgorithmWatch / ITUC, February 2023), https://www.algorithmwatch.org/en/wp-content/uploads/2023/02/2023_AlgorithmWatch_ITUC_Report.pdf
- 94 Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems," *ACM Transactions on Information Systems* 14, no. 3 (1996): 330–347, <https://doi.org/10.1145/230538.230561>
- 95 Giuseppe Desolda, Andrea Esposito, Rosa Lanzilotti, and Antonio Piccinno, "Designing and Evaluating Human-Centred AI Systems: Best-Practices from a Multidisciplinary View," in *Joint Proceedings of the ACM IUI Workshops 2025 (CEUR Workshop Proceedings, Vol. 3957)*, Cagliari, Italy, March 24–27 2025, 170–178, <https://www.ceur-ws.org/Vol-3957/BEHAVEAI-paper05.pdf>
- 96 Arvind Narayanan and Sayash Kapoor, *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference* (Princeton, NJ: Princeton University Press, September 24, 2024), 360 pp, <https://www.aisnakeoil.com/p/starting-reading-the-ai-snake-oil>
- 97 Joe Wilkins, "World's First AI Minister is 'Pregnant' With 83 Offspring Government Announces," *Futurism*, October 28, 2025, <https://futurism.com/artificial-intelligence/rama-diella-albania-pregnant>

- 98 “Congratulations! - EC spox on ‘pregnancy’ of Albanian AI minister, dodges questions on US-China trade talks,” Viory, October 30, 2025, https://www.viory.video/en/videos/a3454_30102025/congratulations-ec-spox-on-pregnancy-of-albanian-ai-minister-dodges-questions-on-us-china-trade-talks
- 99 BBC News, “BBC Verify,” <https://www.bbc.co.uk/news/bbcverify>
- 100 Mariana Mazzucato, *The Entrepreneurial State: Debunking Public vs. Private Sector Myths* (London: Anthem Press, 2013), <https://archive.org/details/entrepreneurials0000mazz>
- 101 Nina Lakhani, “National Science Foundation staff decry Trump’s ‘politically motivated’ cuts,” *The Guardian*, July 24, 2025, <https://www.theguardian.com/us-news/2025/jul/24/national-science-foundation-trumps-cuts>
- 102 National Science & Technology Council, *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update* (Washington, DC: Select Committee on AI, June 21, 2019), <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>
- 103 Montréal Declaration for a Responsible Development of Artificial Intelligence, launched November 3, 2017; official release December 4, 2018, Université de Montréal, <https://montrealdeclaration-responsibleai.com/the-declaration/>

BIBLIOGRAPHY

Abelson, Hal, Ross Anderson, Steven M. Bellovin, et al. “Bugs in Our Pockets: The Risks of Client-Side Scanning”. arXiv, ahead of print, arXiv, 2021. <https://doi.org/10.48550/ARXIV.2110.07450>.

Abrahamoff, Michael D., Philip T. Lavin, Michele Birch, Nilay Shah, and James C. Folk. “Pivotal Trial of an Autonomous AI-Based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices”. *Npj Digital Medicine* 1, no. 1 (2018): 39. <https://doi.org/10.1038/s41746-018-0040-6>.

Acemoglu, Daron, and Pascual Restrepo. “Robots and Jobs: Evidence from US Labor Markets”. *Journal of Political Economy* 128, no. 6 (2020): 2188–244. <https://doi.org/10.1086/705716>.

Ahmed, Nur, Muntasir Wahed, and Neil C. Thompson. “The Growing Influence of Industry in AI Research”. *Science* 379, no. 6635 (2023): 884–86. <https://doi.org/10.1126/science.ade2420>.

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E. Blumenstock. “Machine Learning and Phone Data Can Improve Targeting of Humanitarian Aid”. *Nature* 603, no. 7903 (2022): 864–70. <https://doi.org/10.1038/s41586-022-04484-9>.

Ajder, Henry, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen. “The State of Deepfakes: Landscape, Threats, and Impact”. Docslib, September 2019. <https://docslib.org/doc/12559428/the-state-of-deepfakes-landscape-threats-and-impact-henry-ajder-giorgio-patrini-francesco-cavalli-and-laurence-cullen-september-2019>.

Alexis, Alexei. “AI Startup Reaped Millions Using Bogus Claims, FTC Suit Says”. CFO Dive, 26 August 2025. <https://www.cfodive.com/news/ai-startup-reaped-millions-using-bogus-claims-ftc-suit/758587/>.

AlgorithmWatch gGmbH and International Trade Union Confederation. Algorithmic Transparency and Accountability in the World of Work: A Mapping Study into the Activities of Trade Unions. AlgorithmWatch/ITUC, 2023. https://www.algorithmwatch.org/en/wp-content/uploads/2023/02/2023_AlgorithmWatch_ITUC_Report.pdf.

Algoritmeregister. “Gemeente Amsterdam”. <https://algoritmes.overheid.nl/nl/organisatie/gm0363/gemeente-amsterdam>.

Allyn, Bobby. “Stung By Media Coverage, Silicon Valley Starts Its Own Publications”. NPR, June 25, 2012. <https://www.npr.org/2021/06/25/1010066447/stung-by-media-coverage-silicon-valley-starts-its-own-publications>.

Altman, Sam. “Moore’s Law for Everything”. Moores, 16 March 2021. <https://moores.samaltman.com/>.

Altman, Sam. “Planning for AGI and Beyond”. OpenAI, 24 February 2023. <https://openai.com/index/planning-for-agi-and-beyond/>.

Altman, Sam (@sama). “We Made ChatGPT Pretty Restrictive to Make Sure We Were Being Careful with Mental Health Issues...”. X (Formerly Twitter), 14 October 2025. <https://x.com/sama/status/1978129344598827128>.

Amazon Staff. “Amazon Concludes \$4 Billion Investment in Anthropic”. Amazon, 27 March 2024. <https://www.aboutamazon.com/news/company-news/amazon-anthropic-ai-investment>.

Amodei, Daniela. “Constitutional AI”. Stanford Technology Ventures Program, 23 February 2024. <https://stvp.stanford.edu/clips/constitutional-ai/>.

Amodei, Dario. “Core Views on AI Safety: When, Why, What, and How”. Anthropic, April 2023. <https://www.anthropic.com/news/core-views-on-ai-safety>.

Anderson, Benedict R. O’G. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso, 2006. http://archive.org/details/imaginedcommunit00ande_0.

Anderson, Chris. *TED Talks: The Official TED Guide to Public Speaking*. Headline Publishing Group, 2016. http://archive.org/details/tedtalksofficial0000ande_f3k1.

- Andreessen, Marc. "Why AI Will Save the World". Andreessen Horowitz, 6 June 2023. <https://a16z.com/ai-will-save-the-world/>.
- Andrews, Edmund L. "How Flawed Data Aggravates Inequality in Credit". Stanford HAI, 6 August 2021. <https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>.
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased against Blacks". ProPublica, 23 May 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Anupriya, Datta, and Henning Maximilian. "Irish Election Deepfake Scandal Spotlights Slow Enforcement on AI Fakes". Euractiv, 24 October 2025. <https://www.euractiv.com/news/irish-election-deepfake-scandal-spotlights-slow-enforcement-on-ai-fakes/>.
- Anwer, Afzana. "Economics of Streaming & the Rise of the Music Artists" Rights and Compensation - IP Business Academy. IP Business Academy, 24 February 2025. <https://ipbusinessacademy.org/economics-of-streaming-the-rise-of-the-music-artists-rights-and-compensation>.
- Article 19. Algorithmic People-Pleasers: Are AI Chatbots Telling You What You Want to Hear? 20 May 2025. <https://www.article19.org/resources/algorithmic-people-pleasers-are-ai-chatbots-telling-you-what-you-want-to-hear/>.
- Atlantic Council experts. "Reading Between the Lines of the Dueling US and Chinese AI Action Plans". Atlantic Council, 7 August 2025. <https://www.atlanticcouncil.org/blogs/new-atlanticist/reading-between-the-lines-of-the-dueling-us-and-chinese-ai-action-plans/>.
- Autor, David H. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation". *Journal of Economic Perspectives* 29, no. 3 (2015): 3–30. <https://doi.org/10.1257/jep.29.3.3>.
- Autor, David H., Frank Levy, and Richard J. Murnane. "The Skill Content of Recent Technological Change: An Empirical Exploration". *The Quarterly Journal of Economics* 118, no. 4 (2003): 1279–333. <https://doi.org/10.1162/003355303322552801>.
- Azevedo, Mary Ann. "Microsoft to Spend \$80 Billion in FY'25 on Data Centers for AI". TechCrunch, 3 January 2025. <https://techcrunch.com/2025/01/03/microsoft-to-spend-80-billion-in-fy25-on-data-centers-for-ai/>.

Axios. “Zuckerberg, Chan Launch \$1B Biohub to Build AI to Cure All Diseases”. 6 November 2025. <https://www.axios.com/2025/11/06/zuckerberg-chan-biohub-ai-disease>.

Bagby, John W., and Nizan G. Packin. “Closing the ‘RegLag’ Between Prospective Regulated Activity and Regulation”. CLS Blue Sky Blog, 16 October 2020. <https://clsbluesky.law.columbia.edu/2020/10/16/closing-the-reglag-between-prospective-regulated-activity-and-regulation/>.

Bai, Yuntao, Saurav Kadavath, Sandipan Kundu, et al. “Constitutional AI: Harmlessness from AI Feedback”. arXiv:2212.08073. Preprint, arXiv, 15 December 2022. <https://doi.org/10.48550/arXiv.2212.08073>.

Bank for International Settlements. “CAP10 Definition of Eligible Capital”. Basel Framework, 15 December 2019. https://www.bis.org/basel_framework/chapter/CAP/10.htm.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2023. <https://fairmlbook.org/>.

Barr, Avron, and Shirley Tessler. *Expert Systems: A Technology Before Its Time*. AI Expert, 1995. <http://www.ksl.stanford.edu/people/eaf/cs226/AIExpert95.pdf>.

Baumgartner, Frank R., and Bryan D. Jones. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press, 1993. <http://archive.org/details/agendasinstabili0000baum>.

BBC News. “BBC Verify”. <https://www.bbc.com/news/bbcverify>.

BBC News. “Funeral of Speaking Clock Voice Brian Cobby in Brighton”. Sussex. BBC News, 16 November 2012. <https://www.bbc.com/news/uk-england-sussex-20353154>.

Beach, Thomas H., Yacine Y. Rezgui, Hijiang Li, and T. Kasim. “A Rule-Based Semantic Approach for Automated Regulatory Compliance in the Construction Sector”. *Expert Systems with Applications* 42, no. 12 (2015): 5219–31. <https://doi.org/10.1016/j.eswa.2015.02.029>.

Belga News Agency. “‘We Will Live as One in Heaven’: Belgian Man Dies by Suicide After Chatbot Exchanges”. 28 March 2023. <https://www.belganewsagency.eu/we-will-live-as-one-in-heaven-belgian-man-dies-of-suicide-following-chatbot-exchanges>.

Bellan, Rebecca. “California Becomes First State to Regulate AI Companion Chatbots”. TechCrunch, 13 October 2025. <https://techcrunch.com/2025/10/13/california-becomes-first-state-to-regulate-ai-companion-chatbots/>.

Bender, Emily M. “We’re Going to Need Journalists to Stop Talking About Synthetic Text Extruding Machines...”. LinkedIn, 10 August 2025. https://www.linkedin.com/posts/ebender_were-going-to-need-journalists-to-stop-talking-activity-7359229969345978369-3gnI/.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🐦”. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, ACM, 3 March 2021, 610–23. <https://doi.org/10.1145/3442188.3445922>.

Bender, Emily M., and Alex Hanna. “The AI Con: How to Fight Big Tech’s Hype and Create the Future We Want”. The AI Con. <https://thecon.ai/>.

Bengio, Yoshua. “AI for Social Good”. Yoshua Bengio, 2021. <https://yoshuabengio.org/category/ai-for-social-good/>.

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity Press, 2019.

Benjamin, Walter. “The Work of Art in the Age of Mechanical Reproduction”. In *Illuminations*, edited by Hannah Arendt, translated by Harry Zohn. Schocken Books, 1968. <https://web.mit.edu/allanmc/www/benjamin.pdf>.

Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006.

Berlin, Matthew, and Natasha Weis. “California Enacts Comprehensive AI Safety Law: What AI Developers Need to Know”. ArentFox Schiff, 24 October 2025. <https://www.afslaw.com/perspectives/ai-law-blog/california-enacts-comprehensive-ai-safety-law-what-ai-developers-need-know>.

Bharath, K. “Object Detection Algorithms and Libraries”. Neptune.AI, 22 July 2022. <https://neptune.ai/blog/object-detection-algorithms-and-libraries>.

Bimber, Bruce A. *The Politics of Expertise in Congress: The Rise and Fall of the Office of Technology Assessment*. Albany, NY : State University of New York Press, 1996. <http://archive.org/details/politicsofexpertise0000bimb>.

Black, Julia. *The Rise, Fall and Fate of Principles Based Regulation*. LSE Law, Society and Economy Working Papers 13/2008. London School of Economics and Political Science, 2008. <https://www.lse.ac.uk/law/research/working-paper-series/2007-08/WPS2008-13-Black.pdf>.

Blasi, Weston. “Nvidia Is Now Worth More Than the GDP of Every Country Except These 11”. *Market Watch*, 24 February 2024. <https://www.marketwatch.com/story/nvidia-is-now-worth-more-than-the-gdp-of-every-country-except-these-few-d58a3508>.

Boffey, Daniel. “EU Eyes Temporary Ban on Facial Recognition in Public Places”. *Technology*. *The Guardian*, 17 January 2020. <https://www.theguardian.com/technology/2020/jan/17/eu-eyes-temporary-ban-on-facial-recognition-in-public-places>.

Bond, Shannon. “How AI Deepfakes Polluted Elections in 2024”. *Untangling Disinformation*. NPR, 21 December 2024. <https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections>.

Bordelon, Brendan. “The Law Firm Acting as OpenAI’s Sherpa in Washington”. *POLITICO Pro*, 12 September 2023. <https://subscriber.politicopro.com/article/2023/09/the-law-firm-acting-as-openais-sherpa-in-washington-00115205>.

Borgmann, Albert. *Holding On to Reality: The Nature of Information at the Turn of the Millennium*. University of Chicago Press, 2006. <https://press.uchicago.edu/ucp/books/book/chicago/H/bo3640475.html>.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2016.

Bostrom, Nick, Milan M. Cirkovic, Nick Bostrom, and Milan M. Cirkovic, eds. “Artificial Intelligence as a Positive and Negative Factor in Global Risk”. In *Global Catastrophic Risks*. Oxford University Press, 2011.

Boyd, Danah. “The Politics of ‘Real Names’”. *Communications of the ACM* 55, no. 8 (2012): 29–31. <https://doi.org/10.1145/2240236.2240247>.

Bradford, Anu. *The Brussels Effect: How the European Union Rules the World*. 1st edn. Oxford University PressNew York, 2020. <https://doi.org/10.1093/oso/9780190088583.001.0001>.

Brandom, Russell. “OpenAI completes its for-profit recapitalization”. *TechCrunch*, 28 October 2025. <https://techcrunch.com/2025/10/28/openai-completes-its-for-profit-recapitalization/>.

Brannon, Valerie C., and Eric N. Holmes. “Section 230: An Overview”. *Legislation*. Congress Government. <https://www.congress.gov/crs-product/R46751>.

- Brayne, Sarah. *Predict and Surveil: Data, Discretion, and the Future of Policing*. 1st edn. Oxford University Press, 2021. <https://doi.org/10.1093/oso/9780190684099.001.0001>.
- British Medical Association. *Principles for Artificial Intelligence (AI) and Its Application in Healthcare*. British Medical Association, 2024. <https://www.bma.org.uk/media/njgfbmnn/bma-principles-for-artificial-intelligence-ai-and-its-application-in-healthcare.pdf>.
- Brown, Ellen Hodgson. *The Public Bank Solution: From Austerity to Prosperity*. Baton Rouge, Louisiana : Third Millennium Press, 2013. <http://archive.org/details/publicbanksoluti0000brow>.
- Brugman, Britta C., Christian Burgers, and Barbara Vis. “Metaphorical Framing in Political Discourse Through Words Vs. Concepts: A Meta-Analysis”. *Language and Cognition* 11, no. 1 (2019): 41–65. <https://doi.org/10.1017/langcog.2019.5>.
- Bryson, Joanna. “AI Regulation and the Limits of Transparency”. AXA Research Fund Insight, 12 July 2021. <https://www.axa.com/en/insights/ai-regulation-and-the-limits-of-transparency>.
- Bulatov, Yaroslav. “New Navy Device Learns by Doing”. Medium, 8 October 2024. <https://yaroslavvb.medium.com/new-navy-device-learns-by-doing-62f31d2dc7ba>.
- Burke-Kennedy, Eoin. “Data Centres Now Account for 21% of All Electricity Consumption”. *The Irish Times*, 23 July 2024. <https://www.irishtimes.com/business/2024/07/23/electricity-consumption-by-data-centres-rises-to-21-eclipsing-urban-households/>.
- Burrell, Jenna. “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms”. *Big Data & Society* 3, no. 1 (2016): 2053951715622512. <https://doi.org/10.1177/2053951715622512>.
- Burtell, Matthew, and Helen Toner. “The Surprising Power of Next Word Prediction: Large Language Models Explained, Part 1”. Center for Security and Emerging Technology, 8 March 2024. <https://cset.georgetown.edu/article/the-surprising-power-of-next-word-prediction-large-language-models-explained-part-1/>.
- Caballar, Rina Diani, and Cole Stryker. “What Is Computer Vision?” IBM, 17 September 2025. <https://www.ibm.com/think/topics/computer-vision>.
- Caddy, Becca. “I Stopped Saying Thanks to ChatGPT – Here’s What Happened”. TechRadar, 3 March 2025. <https://www.techradar.com/computing/artificial-intelligence/i-stopped-saying-thanks-to-chatgpt-heres-what-happened>.

Cai, Kenrick. “Google Hires Top Talent from Startup Character. AI, Signs Licensing Deal”. *Artificial Intelligence*. Reuters, 2 August 2024. <https://www.reuters.com/technology/artificial-intelligence/google-hires-characterai-cofounders-licenses-its-models-information-reports-2024-08-02/>.

California Legislature. “Assembly Bill No. 730 (2019–2020 Reg. Sess.): Elections: Deceptive Audio or Visual Media”. 10 April 2019. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730.

California Legislature. “California SB 53, Artificial Intelligence Models: Large Developers, 2025-2026 Reg. Sess. (Cal. 2025)”. https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB53.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. “Semantics Derived Automatically from Language Corpora Contain Human-Like Biases”. *Science* 356, no. 6334 (2017): 183–86. <https://doi.org/10.1126/science.aal4230>.

Calo, Ryan. “Products Liability Law as a Way to Address AI Harms”. Brookings, 31 October 2019. <https://www.brookings.edu/articles/products-liability-law-as-a-way-to-address-ai-harms/>.

Calo, Ryan. “Robotics and the Lessons of Cyberlaw”. SSRN Scholarly Paper No. 2402972. Social Science Research Network, 28 February 2014. <https://doi.org/10.2139/ssrn.2402972>.

Calvert, Clay. “Regulating Cyberspace: Metaphor, Rhetoric, Reality, and the Framing of Legal Options”. *UC Law SF Communications and Entertainment Journal* 20, no. 3 (1998): 541.

Carlini, Nicholas, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Florian Tramer, and

Carpio, Alvin. “The Rise of the Political Entrepreneur and Why We Need More of Them”. World Economic Forum, 23 November 2017. <https://www.weforum.org/stories/2017/11/the-rise-of-the-political-entrepreneur-and-why-we-need-more-of-them/>.

Chiyuan Zhang. “Quantifying Memorization Across Neural Language Models”. Version 3. Preprint, arXiv, 2022. <https://doi.org/10.48550/ARXIV.2202.07646>.

Carlini, Nicholas, Florian Tramer, Eric Wallace, et al. “Extracting Training Data from Large Language Models”. arXiv:2012.07805. Preprint, arXiv, 15 June 2021. <https://doi.org/10.48550/arXiv.2012.07805>.

Carvalho, Julio. "The Stubborn Memory of Generative AI: Overfitting, Fair Use, and the AI Act". Kluwer Copyright Blog, 8 April 2024. <https://legalblogs.wolterskluwer.com/copyright-blog/the-stubborn-memory-of-generative-ai-overfitting-fair-use-and-the-ai-act/>.

Cave, Stephen, Claire Craig, Kanta Dihal, et al. *Portrayals and Perceptions of AI and Why They Matter*. With Apollo-University Of Cambridge Repository and University Of Cambridge. The Royal Society, 2018. <https://doi.org/10.17863/CAM.34502>.

Ceballos Ferroglio, Carlos F., Gustavo Ferro, and Ángel Enrique Neder. "Regulatory Lag, Efficiency, and Performance. Lessons from a Case Study". *Competition and Regulation in Network Industries* 25, no. 1 (2024): 43–66. <https://doi.org/10.1177/17835917241251811>.

Center for AI Safety. "Statement on AI Risk". 30 May 2023. <https://aistatement.com>.

Centre for Trustworthy Technology. "Metaphors Matter: How Language Influences Technology and Society". Uncategorized. Centre for Trustworthy Technology, 17 June 2024. <https://c4tt.org/metaphors-matter-how-language-influences-technology-and-society/>.

Cerf, Vinton G. "Our 21st Century Information Superhighway (Adapted From The Print Edition of Ingenia No. 82)". *Ingenia*, March 2020. <https://www.ingenia.org.uk/articles/our-21st-century-information-superhighway/>.

Charnock, Greig, Hug March, and Ramon Ribera-Fumaz. "From Smart to Rebel City? Worlding, Provincialising and the Barcelona Model". *Urban Studies* 58, no. 3 (2021): 581–600. <https://doi.org/10.1177/0042098019872119>.

Cheney-Lippold, John. *We Are Data: Algorithms and the Making of Our Digital Selves*. NYU Press, 2017. <https://doi.org/10.2307/j.ctt1gk0941>.

Cheng, Myra, Angela Y. Lee, Kristina Rapuano, Kate Niederhoffer, Alex Liebscher, and Jeffrey Hancock. "From Tools to Thieves: Measuring and Understanding Public Perceptions of AI through Crowdsourced Metaphors". *arXiv*, 31 January 2025. <https://arxiv.org/html/2501.18045v1>.

Chesney, Robert, and Danielle Keats Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security". SSRN Scholarly Paper No. 3213954. Social Science Research Network, 14 July 2018. <https://doi.org/10.2139/ssrn.3213954>.

Chiu, Matthew. "Path Dependency in Policymaking: A Double-Edged Sword". Protopia Group, 19 April 2024. <https://www.protopiagroup.org/analysis/path-dependency-in-policymaking-a-double-edged-sword>.

Christensen, Clayton M. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press, 1997. <http://archive.org/details/innovatorsdilem000chri>.

City of Helsinki. "Ethical Principles and the Algorithm Register". Helsinki Data Strategy, 2 June 2025. <https://www.hel.fi/en/decision-making/information-on-helsinki/design-and-digitalisation/digital-helsinki>.

Clarke, Richard A., and Robert K. Knake. "Cyber War: The Next Threat to National Security and What to Do About It". HarperCollins. <https://www.harpercollins.com/products/cyber-war-richard-a-clarkerobert-knake?variant=32123019100194>.

CNBC. "Putin: Leader in Artificial Intelligence Will Rule World". 4 September 2017. <https://www.cnn.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html>.

Coglianese, Cary, and David Lehr. "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era". *Georgetown Law Journal* 105, no. 5 (2017): 1147–223.

Cohen, Julie E. "The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy". SSRN Scholarly Paper No. 2666570. Social Science Research Network, 28 September 2015. <https://doi.org/10.2139/ssrn.2666570>.

Cohen, Stanley, *Folk Devils, and Moral Panics*. *Folk Devils and Moral Panics*. MacGibbon and Kee, 1972. <https://www.routledge.com/Folk-Devils-and-Moral-Panics/Cohen/p/book/9780415610162>.

Cole, Samantha. "'My AI Is Sexually Harassing Me': Repika Users Say the Chatbot Has Gotten Way Too Horny". VICE, 12 January 2023. <https://www.vice.com/en/article/my-ai-is-sexually-harassing-me-repika-chatbot-nudes/>.

Coleman, James S. "Social Capital in the Creation of Human Capital". *American Journal of Sociology* 94, no. 1 (1988): S95–120.

Collins, Jim. "Productive Paranoia". 2025. <https://www.jimcollins.com/concepts/productive-paranoia.html>.

Committee on the Judiciary and Select Subcommittee on the Weaponization of the Federal Government. *The Weaponization of CISA: How a "Cybersecurity" Agency Colluded with Big Tech and "Disinformation" Partners to Censor Americans*. U.S. House of

Representatives, 2023. https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/cisa-staff-report6-26-23.pdf?utm_source=chatgpt.com.

Corder, Mike, and Molly Quell. "Wilders and Timmermans Are Among the Key Political Party Leaders in the Dutch Election". 10tv.com, 28 October 2025. <https://www.10tv.com/article/syndication/associatedpress/wilders-timmermans-are-among-the-leaders-of-the-key-parties-in-dutch-election/616-5c83cae5-d1d0-4db6-9f08-e6435ce1d6c8>.

Craig, Carys J. "The AI-Copyright Trap". SSRN Scholarly Paper No. 4905118. Social Science Research Network, 15 July 2024. <https://doi.org/10.2139/ssrn.4905118>.

Crane, Diana. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. University of Chicago Press, 1972. <http://archive.org/details/invisiblecollege0000cran>.

Dahl, Robert A. *Democracy and Its Critics*. Yale University Press, 1989.

Dario Amodèi. "The Urgency of Interpretability". April 2025. <https://www.darioamodei.com/post/the-urgency-of-interpretability>.

Dartmouth. "Artificial Intelligence (AI) Coined at Dartmouth". Trustees of Dartmouth College. <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth>.

Dastin, Jeffrey. "Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women". World. Reuters, 11 October 2018. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>.

Davis, Will. "Eisenhower's Atomic Power for Peace – The Civilian Application Program". ANS Nuclear Cafe, 20 March 2014. <https://www.ans.org/news/article-1486/atomic-power-for-peace-the-civilian-application-program-and-power-demonstration-reactor/>.

De Cock, Caroline. "Copyright, Creativity, and Generative AI: The Research Sector's Missing Seat at the Policy Table". Information Labs, 11 June 2025. <https://informationlabs.org/copyright-creativity-and-generative-ai-the-research-sectors-missing-seat-at-the-policy-table/>.

De Cock, Caroline. «Digital Sovereignty: Europe's Favourite (Empty) Buzzword». Information Labs, 22 October 2025. <https://informationlabs.org/digital-sovereignty-europes-favourite-empty-buzzword/>.

Debenedetti, Luciana. “Togo’s Novissi Cash Transfer: Designing and Implementing a Fully Digital Social Assistance Program during COVID-19”. *Innovations for Poverty Action*, 4 August 2021. <https://poverty-action.org/publication/togo%E2%80%99s-novissi-cash-transfer-designing-and-implementing-fully-digital-social-assistance>.

Decidim, “Decidim is a digital platform for citizen participation”. <https://decidim.org/>.

Decidim Association. “Decidim: Free OpenSource Participatory Democracy Platform”. 2024. <https://doi.org/10.1007/978-3-031-50784-7>.

Defense Advanced Research Projects Agency. “MediFor: Media Forensics”. <https://www.darpa.mil/research/programs/media-forensics>.

Defense Advanced Research Projects Agency. “SemaFor: Semantic Forensics”. DARPA. <https://www.darpa.mil/research/programs/semantic-forensics>.

De Ruvo, Giuseppe. “Algorithmic Objectivity as Ideology: Toward a Critical Ethics of Digital Capitalism”. *Topoi* 44 (March 2025): 175–86. <https://doi.org/10.1007/s11245-024-10117-9>.

Desolda, Giuseppe, Andrea Esposito, Rosa Lanzilotti, and Antonio Piccinno. “Designing and Evaluating Human-Centred AI Systems: Best-Practices from a Multidisciplinary View”. *Joint Proceedings of the ACM IUI Workshops 2025 (CEUR Workshop Proceedings 3957 (March 2025))*: 170–78.

Devine, Curt, Donie O’Sullivan, and Sean Lyngaas. “A Fake Recording of a Candidate Saying He’d Rigged the Election Went Viral. Experts Say It’s Only the Beginning”. *CNN*, 1 February 2024. <https://www.cnn.com/2024/02/01/politics/election-deepfake-threats-invs>.

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market (“Directive on Electronic Commerce”), CONSIL, EP, 178 OJ L (2000). <http://data.europa.eu/eli/dir/2000/31/oj>.

Doctorow, Cory. “Don’t Believe the Criti-Hype”. *Pluralistic*, 18 December 2022. <https://pluralistic.net/2021/09/30/dont-believe-the-criti-hype/>.

Doctorow, Cory. “Pluralistic: “Privacy Preserving Age Verification” Is Bullshit”. *Pluralistic*, 16 September 2025. <https://pluralistic.net/2025/08/14/bellovin/>.

- Donovan, Joan. *Meme Wars: The Untold Story of the Online Battles Upending Democracy in America*. Bloomsbury, 2022. <https://www.bloomsbury.com/us/meme-wars-9781635578638/>.
- Dworkin, Ronald. *Law's Empire*. Belknap Press of Harvard University Press, 1986. <http://archive.org/details/lawsempire0000dwor>.
- Echikson, William. "Limited Liability for the Net? The Future of Europe's E-Commerce Directive". CEPS, 13 April 2017. <https://www.ceps.eu/ceps-publications/limited-liability-net-future-europes-e-commerce-directive/>.
- Edelman, Murray J. *Constructing the Political Spectacle*. Chicago : University of Chicago Press, 1988. <http://archive.org/details/constructingpoli00edel>.
- Edwards, Paul N. *The Closed World: Computers and the Politics of Discourse in Cold War America*. The MIT Press, 1996. <https://doi.org/10.7551/mitpress/1871.001.0001>.
- Electronic Frontier Foundation. "Age Verification Mandates Would Undermine Anonymous Speech and Require Extensive Data Collection". March 2023. <https://www.eff.org/deeplinks/2023/03/age-verification-mandates-would-undermine-anonymous-speech-and-require-extensive>.
- Electronic Frontier Foundation. "Section 230". <https://www.eff.org/issues/cda230>.
- "Elon Musk Interviewed at MIT Centennial Celebration, October 2014". Space Policy Online, 24 October 2014. <https://spacepolicyonline.com/meeting-summaries/elon-musk-interviewed-at-mit-centennial-celebration-october-2014/>.
- Emanuilov, Ivon, and Thomas Margoni. "Memorisation in Generative Models and EU Copyright Law: An Interdisciplinary View". Kluwer Copyright Blog, 26 March 2024. <https://legalblogs.wolterskluwer.com/copyright-blog/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/>.
- Environmental Protection Agency. "Toxics Release Inventory (TRI) Program". Collections and Lists. U.S. Environmental Protection Agency, 31 January 2013. <https://www.epa.gov/toxics-release-inventory-tri-program>.
- Estrada, Sheryl. "MIT Report: 95% of Generative AI Pilots at Companies Are Failing". Fortune, 18 August 2025. <https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/>.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018. <https://us.macmillan.com/books/9781250074317/automatinginequality/>.

European Parliament. "EU AI Act: First Regulation on Artificial Intelligence". 6 August 2023. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.

European Parliament Committee on Legal Affairs (JURI). Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (PE 582.443v0100, JURIPR 582443). 2015/2103(INL). European Parliament, 2016. https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf.

European Parliamentary Research Service. "Digital Sovereignty for Europe". July 2020. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI\(2020\)651992_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf).

European Parliament. "Resolution on Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters". 6 October 2021. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html.

European Commission. "Implementation Timeline". EU Artificial Intelligence Act, n.d. <https://artificialintelligenceact.eu/implementation-timeline/>.

Evans, Brad, and Chantal Meza. "The Techno-Theodicy: How Technology Became the New Religion". *Theory & Event* 25, no. 3 (2022): 639–64. <https://doi.org/10.1353/tae.2022.0031>.

Executive Office of the President, Office of Science and Technology Policy. *Winning the Race: America's AI Action Plan*. The White House, 2025. <https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf>.

Farrell, David M., Jane Suiter, and Clodagh Harris. "'Systematizing' Constitutional Deliberation: The 2016–18 Citizens' Assembly in Ireland". *Irish Political Studies* 34, no. 1 (2019): 113–23. <https://doi.org/10.1080/07907184.2018.1534832>.

Farrell, James. "BuzzFeed's Stock Price Soars as Company Announces AI Content Push". *AI. SiliconANGLE*, 26 January 2023. <https://siliconangle.com/2023/01/26/buzzfeeds-stock-price-soars-company-announces-ai-content-push/>.

Feffer, Michael, Anusha Sinha, Wesley Hanwen Deng, Zachary C. Lipton, and Hoda Heidari. "Red-Teaming for Generative AI: Silver

- Bullet or Security Theater?” arXiv:2401.15897. Preprint, arXiv, 27 August 2024. <https://doi.org/10.48550/arXiv.2401.15897>.
- Floridi, Luciano, and Anna C Nobre. “Anthropomorphising Machines and Computerising Minds: The Crosswiring of Languages between Artificial Intelligence and Brain & Cognitive Sciences”. *Minds and Machines* 34, no. 1 (2024): 5–14. <https://doi.org/10.1007/s11023-024-09670-4>.
- Frank, Jerome. *Courts on Trial: Myth and Reality in American Justice*. Princeton University Press, 1949.
- Frankel, Jeffrey A. “Over-Optimism in Forecasts by Official Budget Agencies and Its Implications”. *Oxford Review of Economic Policy* 27, no. 4 (2011): 536–62.
- Frazier, Kevin, and Andrew W. Reddie. “Why Context, Not Compute, Is the Key to AI Governance”. *Tech Policy Press*, 5 August 2025. <https://techpolicy.press/why-context-not-compute-is-the-key-to-ai-governance>.
- French Government. *Make France an AI Powerhouse*. 2025. <https://www.elysee.fr/admin/upload/default/0001/17/d9c1462e7337d353f918aac7d654b896b77c5349.pdf>.
- Frey, Carl Benedikt, and Michael A. Osborne. “The Future of Employment: How Susceptible Are Jobs to Computerisation?” *Technological Forecasting and Social Change* 114, no. 1 (2017): 254–80. <https://doi.org/10.1016/j.techfore.2016.08.019>.
- Friedman, Batya, Peter H. Kahn, Alan Borning, and Alina Hultgren. “Value Sensitive Design and Information Systems”. In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman. Springer Netherlands, 2013. https://doi.org/10.1007/978-94-007-7844-3_4.
- Friedman, Batya, and Helen Nissenbaum. “Bias in Computer Systems”. *ACM Transactions on Information Systems* 14, no. 3 (1996): 330–47. <https://doi.org/10.1145/230538.230561>.
- Froman, Michael. “China, the United States, and the AI Race”. *Council on Foreign Relations*, 24 October 2025. <https://www.cfr.org/article/china-united-states-and-ai-race>.
- Furze, Leon. “AI Metaphors We Live By: The Language of Artificial Intelligence”. *Leon Furze*, 19 July 2024. <https://leonfurze.com/2024/07/19/ai-metaphors-we-live-by-the-language-of-artificial-intelligence/>.

Furze, Leon. "Myths, Magic, and Metaphors: The Language of Generative AI". Leon Furze, 17 October 2023. <https://leonfurze.com/2023/10/18/myths-magic-and-metaphors-the-language-of-generative-ai/>.

Future of Life Institute. "Pause Giant AI Experiments: An Open Letter". Future of Life Institute, 22 March 2023. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

Galbi, Douglas. *The Creation of the Media: Political Origins of Modern Communications*. EH.NET, 2004. https://eh.net/book_reviews/the-creation-of-the-media-political-origins-of-modern-communications/.

Gallagher, Leigh. *The Airbnb Story: How Three Ordinary Guys Disrupted an Industry, Made Billions... and Created Plenty of Controversy*. Houghton Mifflin Harcourt, 2017. <http://archive.org/details/airbnbstoryhowth0000gall>.

Gans, Herbert J. *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time* / Herbert J. Gans. Northwestern University Press, 2004.

Gao, Leo, Stella Biderman, Sid Black, et al. "The Pile: An 800GB Dataset of Diverse Text for Language Modeling". Version 1. Preprint, arXiv, 2021. <https://doi.org/10.48550/ARXIV.2101.00027>.

Gertz, Geoffrey, and Justin Sherman. "Beyond Bans: Expanding the Policy Options for Tech-Security Threats". *Lawfare*, Lawfare, 25 June 2025. <https://www.lawfaremedia.org/article/beyond-bans--expanding-the-policy-options-for-tech-security-threats>.

Getty Images (US) Inc and others v Stability AI Ltd [2025] EWHC 2863 (Ch). <https://www.judiciary.uk/judgments/getty-images-v-stability-ai/>.

Gibson, Lydialyle. "Bias in Artificial Intelligence". *Harvard Magazine*, 2 August 2021. <https://www.harvardmagazine.com/2021/08/meredith-broussard-ai-bias-documentary>.

Giebel, Godwin Denk, Pascal Raszke, Hartmuth Nowak, et al. "Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study". *Journal of Medical Internet Research* 27 (February 2025): e63377. <https://doi.org/10.2196/63377>.

Gilardi, Fabrizio, and Fabio Wasserfallen. "The Politics of Policy Diffusion". *European Journal of Political Research* 58, no. 4 (2019): 1245–56. <https://doi.org/10.1111/1475-6765.12326>.

Gilens, Martin, and Benjamin I. Page. "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens". *Perspectives on Politics* 12, no. 3 (2014): 564–81. <https://doi.org/10.1017/S1537592714001595>.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, 2018. http://archive.org/details/Custodians_of_the_Internet_by_Tarleton_Gillespie.

Global Partnership on Artificial Intelligence. "The OECD Artificial Intelligence Policy Observatory". <https://oecd.ai/en/>.

Gonçalves, Andreu Belsunces, and Jascha Bareis. "Expanding Hype Literacy to Protect Democracy". Tech Policy Press, 21 August 2025. <https://techpolicy.press/expanding-hype-literacy-to-protect-democracy>.

Goode, Lauren. "Welcome to the Age of Paranoia as Deepfakes and Scams Abound". *Ars Technica*, 13 May 2025. <https://www.wired.com/story/paranoia-social-engineering-real-fake/>.

Google DeepMind. "AlphaFold: A Solution to a 50-Year-Old Grand Challenge in Biology". Google DeepMind, January 2022. <https://deepmind.google/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/>.

Goujard, Clothilde. "Italian Privacy Regulator Bans ChatGPT". *Politico*, 31 March 2023. <https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/>.

Goujard, Clothilde. "Italian Privacy Regulator Bans ChatGPT". *Politico*. <https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/>.

Government of Canada. "Algorithmic Impact Assessment Tool". 30 May 2024. <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>.

GovTech News Staff. "New York City Department of Education Bans ChatGPT". GovTech, 10 January 2023. <https://www.govtech.com/education/k-12/new-york-city-department-of-education-bans-chatgpt>.

Gray, Mary L., and Siddharth Suri. "Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass (Co-Authoring with Siddharth Suri)". Houghton Mifflin Harcourt, 7 May 2019. <https://www.microsoft.com/en-us/research/publication/ghost-work-how-to-stop-silicon-valley-from-building-a-new-global-underclass-co-authored-with-siddharth-suri/>.

Griffin, Andrew. “Facebook Robots Shut down After They Talk to Each Other in Language Only They Understand”. *The Independent*, 10 September 2020. <https://www.independent.co.uk/life-style/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html>.

Guadamuz, Andres. “Artists File Class-Action Lawsuit Against Stability AI, DeviantArt, and Midjourney”. *TechnoLlama*, 15 January 2023. <https://www.technollama.co.uk/artists-file-class-action-lawsuit-against-stability-ai-deviantart-and-midjourney>.

Guadamuz, Andres. “Getty Images v Stability AI: A landmark High Court ruling on AI, copyright, and trade marks”. *Technollama*, 6 November 2025. <https://www.technollama.co.uk/getty-images-v-stability-ai-a-landmark-high-court-ruling-on-ai-copyright-and-trade-marks>.

Guadamuz, Andres. “How AI Is Breaking Traditional Remuneration Models”. *TechnoLlama*, 9 July 2025. <https://www.technollama.co.uk/how-ai-is-breaking-traditional-remuneration-models>.

Habgood-Coote, Joshua. “Deepfakes and the Epistemic Apocalypse”. *Synthese* 201, no. 3 (2023): 103. <https://doi.org/10.1007/s11229-023-04097-3>.

Hall, Peter A. “Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain”. *Comparative Politics* 25, no. 3 (1993): 275. <https://doi.org/10.2307/422246>.

Halm, Kevin Carl, Jhon D Seiver, and Patrick J Austin. “Italy’s Data Protection Agency Lifts Ban on ChatGPT”. *Davis Wright Tremaine*, 15 May 2023. <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/05/ai-chatgpt-italy-ban-lifted>.

Hanna, Alex, and Emily M. Bender. “AI Causes Real Harm. Let’s Focus on That over the End-of-Humanity Hype”. *Scientific American*, 12 August 2023. <https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/>.

Hao, Karen. *Empire of AI*. Allen Lane, 2025. <https://www.penguin.co.uk/books/460331/empire-of-ai-by-hao-karen/9780241678923>.

Hartzog, Woodrow. “Review of Privacy’s Blueprint: The Battle to Control the Design of New Technologies”. *Leonardo/ISAST*, 29 April 2021. <https://leonardo.info/review/2018/12/review-of-privacy-blueprint-the-battle-to-control-the-design-of-new-technologies>.

Hector, Hamish. “Are You Polite to ChatGPT? Here’s Where You Rank Among AI Chatbot Users”. *TechRadar*, 20 February 2025.

<https://www.techradar.com/computing/artificial-intelligence/are-you-polite-to-chatgpt-heres-where-you-rank-among-ai-chatbot-users>.

Heikkilä, Melissa. "Dutch scandal serves as a warning for Europe over risks of using algorithms". POLITICO, 29 March 2022. <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>.

Heikkilä, Melissa. "The 'Hallucinations' That Haunt AI: Why Chatbots Struggle to Tell the Truth". Financial Times, 22 July 2025. <https://www.ft.com/content/7a4e7eae-f004-486a-987f-4a2e4dbd34fb>.

Heikkilä, Melissa. "Why It'll Be Hard to Tell If AI Ever Becomes Conscious". MIT Technology Review, 17 October 2023. <https://www.technologyreview.com/2023/10/17/1081818/why-itll-be-hard-to-tell-if-ai-ever-becomes-conscious/>.

Henley, Jon. "Albania puts AI-created 'minister' in charge of public procurement". The Guardian, 11 September 2025. <https://www.theguardian.com/world/2025/sep/11/albania-diella-ai-minister-public-procurement>.

Hern, Alex. "Give Robots 'Personhood' Status, EU Committee Argues". Technology. The Guardian, 12 January 2017. <https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues>.

Heschel, Abraham Joshua. *Moral Grandeur and Spiritual Audacity: Essays*. Edited by Susannah Heschel. Farrar, Straus & Giroux, 1997. <http://archive.org/details/moralgrandeurspi0000hesc>.

Hierlemann, Dr. Dominik, and Dr. Stefan Roch. *Digital Democracy – What Europe Can Learn from Taiwan*. Bertelsmann Stiftung, 2020. https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Demokratie_und_Partizipation_in_Europa_/210211_Bericht_Digital_Democracy_-_What_Europe_can_learn_from_Taiwan__1_.pdf.

Hill, Kashmir. "A Teen Was Suicidal. ChatGPT Was the Friend He Confided In." Technology. The New York Times, 26 August 2025. <https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html>.

Hindman, Matthew. *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy*. 1st edn. Princeton University Press, 2018. <https://doi.org/10.23943/princeton/9780691159263.001.0001>.

Hlomani, Hanani, Mark Gaffley, Liz Orembo, Theresa Schültken, and Scott Timcke. "Dissent and Resistance to Silicon Valley AI

Narratives". Tech Policy Press, 21 March 2023. <https://techpolicy.press/dissent-and-resistance-to-silicon-valley-ai-narratives>.

Horsey, Julian. "Are You Talking to Bots? The Alarming Truth About the Dead Internet Theory". Geeky Gadgets, 1 July 2025. <https://www.geeky-gadgets.com/dead-internet-theory-explained/>.

Horton, Michael J. M. "Artificial Intelligence & Product Liability". McCarter & English, LLP, 20 August 2024. <https://www.mccarter.com/insights/artificial-intelligence-product-liability/>.

Horwitz, Jeff. "Meta Let Chatbots Engage in Inappropriate Conversations with Minors, Reuters Reports". Business & Human Rights Resource Centre, 14 August 2025. <https://www.business-humanrights.org/en/latest-news/meta-let-chatbots-engage-in-inappropriate-conversations-with-minors-reuters-reports/>.

"House of Lords - Regulating in a Digital World - Select Committee on Communications". <https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/29908.htm>.

House Permanent Select Committee on Intelligence. "National Security Challenges of Artificial Intelligence, Manipulated Media, and 'Deepfakes'". Legislation. Congress Government, 13 June 2019. <https://www.congress.gov/event/116th-congress/house-event/109620>.

"How AI Is Learning to Think on Its Own Like Humans". Technology. SciTechDaily, 18 September 2024. <https://scitechdaily.com/how-ai-is-learning-to-think-on-its-own-like-humans/>.

Hsiao, Yu T, Shu-Yang Lin, Audrey Tang, Darshana Narayanan, and Claudina Sarahe. "vTaiwan: An Empirical Study of Open Consultation Process in Taiwan". Xyhft_v1. Preprint, SocArXiv, 4 July 2018. <https://doi.org/10.31235/osf.io/xyhft>.

Hudson, Bob, David Hunter, and Stephen Peckham. "Policy Failure and the Policy-Implementation Gap: Can Policy Support Programs Help?" *Policy Design and Practice* 2, no. 1 (2019): 1–14. <https://doi.org/10.1080/25741292.2018.1540378>.

Huszár, Ferenc, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. "Algorithmic Amplification of Politics on Twitter". *Proceedings of the National Academy of Sciences* 119, no. 1 (2022): e2025334119. <https://doi.org/10.1073/pnas.2025334119>.

Hutson, Matthew. "AI Researchers Allege That Machine Learning Is Alchemy". *Science*, 3 May 2018. <https://www.science.org/content/article/ai-researchers-allege-machine-learning-alchemy>.

“Imagining the Internet | Elon University. “1830s – 1860s: Telegraph”. <https://www.elon.edu/u/imagining/time-capsule/150-years/back-1830-1860/>.

“International Civil Aviation Organization”. <https://www.icao.int/>.

“International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)”. <https://www.ich.org/>.

International Energy Agency (IEA). Energy and AI. IEA, 2025. <https://www.iea.org/reports/energy-and-ai>.

Investopedia. “Understanding Wire Fraud: Definition, Laws, and Key Examples”. Investopedia, 27 August 2025. <https://www.investopedia.com/terms/w/wirefraud.asp>.

Irish Council for Civil Liberties. “Scientists call on the President of the European Commission to retract AI hype statement.” November 10, 2025. <https://www.iccl.ie/press-release/scientists-call-on-the-president-of-the-european-commission-to-retract-ai-hype-statement/>.

Jasanoff, Sheila, ed. *States of Knowledge: The CoProduction of Science and the Social Order*. 0 edn. Routledge, 2004. <https://doi.org/10.4324/9780203413845>.

Jensen, Casper Bruun. “Citizen Projects and Consensus-Building at the Danish Board of Technology: On Experiments in Democracy”. *Acta Sociologica* 48, no. 3 (2005): 221–35.

Jureta, Ivan. “Policy Windows: What They Are And When They Occur”. Ivan Jureta blog, December 2024. <https://ivanjureta.com/policy-windows-what-they-are-and-when-they-occur/>.

Kahle, Brewster. “Imagining the Internet: Explaining Our Digital Transition”. Brewster Kahle’s Blog, 13 January 2022. <https://brewster.kahle.org/2022/01/13/imagining-the-internet-explaining-our-digital-transition/>.

Kak, Amba, and Sarah West. “2. A Modern Industrial Strategy for AI?: Interrogating the US Approach”. AI Now Institute, 12 March 2024. <https://ainowinstitute.org/publications/a-modern-industrial-strategy-for-aiinterrogating-the-us-approach>.

Kalvet, Tarmo. “Innovation: A Factor Explaining E-Government Success in Estonia”. *Electronic Government, an International Journal* 9, no. 2 (2012): 142. <https://doi.org/10.1504/EG.2012.046266>.

Kanno-Youngs, Zolan, Madeleine Ngo, and Don Clark. “Intel Receives \$8.5 Billion in Grants to Build Chip Plants”. U.S. The New

York Times, 20 March 2024. <https://www.nytimes.com/2024/03/20/us/politics/chips-act-grant-intel.html>.

Kaplan, Jared, Sam McCandlish, Tom Henighan, et al. “Scaling Laws for Neural Language Models”. Version 1. Preprint, arXiv, 2020. <https://doi.org/10.48550/ARXIV.2001.08361>.

Keen, Braeden. “Principles-Based Regulations: What Are the Opportunities and Trade-Offs?” Australian Energy Council, 8 May 2025. <https://www.energycouncil.com.au/analysis/principles-based-regulations-what-are-the-opportunities-and-trade-offs/>.

Kellner, Douglas, and Jeff Share. “Toward Critical Media Literacy: Core Concepts, Debates, Organizations, and Policy”. *Discourse: Studies in the Cultural Politics of Education* 26, no. 3 (2005): 369–86. <https://doi.org/10.1080/01596300500200169>.

Kelly, Jack. “Goldman Sachs Predicts 300 Million Jobs Will Be Lost Or Degraded By Artificial Intelligence”. *Forbes*, 31 March 2023. <https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/>.

Kerwin, Cornelius M, and Scott R Furlong. *Rulemaking: How Government Agencies Write Law and Make Policy*. 4th edn. CQ Press, 2010. http://archive.org/details/rulemakinghowgov0000kerw_4edi.

Kestin, Gregory, Kelly Miller, Anna Klaes, Timothy Milbourne, and Gregorio Ponti. “AI Tutoring Outperforms Active Learning”. Preprint, In Review, 14 May 2024. <https://doi.org/10.21203/rs.3.rs-4243877/v1>.

Khan, Lina. “Amazon’s Antitrust Paradox”. SSRN Scholarly Paper No. 2911742. Social Science Research Network, 31 January 2017. <https://papers.ssrn.com/abstract=2911742>.

Khanal, Shaleen, Hongzhou Zhang, and Araz Tacihagh. “Why and How Is the Power of Big Tech Increasing in the Policy Process? The Case of Generative AI”. *Policy and Society* 44, no. 1 (2025): 52–69. <https://doi.org/10.1093/polsoc/puae012>.

Kiese Bahangulu, Julien and Louis Owusu-Berko. “Algorithmic Bias, Data Ethics, and Governance: Ensuring Fairness, Transparency and Compliance in AI-Powered Business Analytics Applications”. *World Journal of Advanced Research and Reviews* 25, no. 2 (2025): 1746–63. <https://doi.org/10.30574/wjarr.2025.25.2.0571>.

Kim, Pauline. “Data-Driven Discrimination at Work”. *William & Mary Law Review* 58, no. 3 (2017): 857.

- Kingdon, John W. *Agendas, Alternatives, and Public Policies*. 2nd edn. HarperCollinsCollege, 1995. <http://archive.org/details/agendasalternati00king>.
- Kitchin, Rob. *The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures*. SAGE Publications Ltd, 2014.
- Klein, Elana. “Tech in the Classroom: A History of Hype and Hysteria”. Tags. *Wired*, 18 August 2025. <https://www.wired.com/story/photo-essay-school-tech-hysteria/>.
- Klein, Julie Thompson. *Interdisciplinarity: History, Theory, and Practice*. Wayne State University Press, 1990. <http://archive.org/details/interdisciplinar00kleirich>.
- Knight, Will. “Many Top AI Researchers Get Financial Backing From Big Tech”. Tags. *Wired*, 4 October 2020. <https://www.wired.com/story/top-ai-researchers-financial-backing-big-tech/>.
- Koebler, Jason. “Is Google’s AI Actually Discovering ‘Millions of New Materials?’” 404 Media, 11 April 2024. <https://www.404media.co/google-says-it-discovered-millions-of-new-materials-with-ai-human-researchers/>.
- Kosseff, Jeff. *The Twenty-Six Words That Created the Internet* by Jeff Kosseff. Cornell University Press, 2019. <https://www.cornellpress.cornell.edu/book/9781501714412/the-twenty-six-words-that-created-the-internet/>.
- Kubala, Szymon. “AI and the Dead Internet Theory: Are We Heading in That Direction?” WebMakers Software House, 20 August 2024. <https://webmakers.expert/en/blog/ai-and-the-dead-internet-theory>.
- Kulesz, Octavio. *Artificial Intelligence and International Cultural Relations: Challenges and Opportunities for Cross-Sector Collaboration*. IFA, 2024. https://opus.bsz-bw.de/ifa/frontdoor/deliver/index/docId/1203/file/ifa-2024_kulesz_ai-intl-cultural-relations.pdf.
- Kundu, Oishee, Elvira Uyarra, Raquel Ortega-Argiles, Mayra M Tirado, Tasos Kitsos, and Pei-Yu Yuan. “Impacts of Policy-Driven Public Procurement: A Methodological Review”. *Science and Public Policy* 52, no. 1 (2025): 50–64. <https://doi.org/10.1093/scipol/scae058>.
- Kurzweil, Ray. *The Singularity Is Near by: When Humans Transcend Biology*. Penguin Random House Books, 2005. <https://www.penguinrandomhouse.com/books/291221/the-singularity-is-near-by-ray-kurzweil/>.

Kyprianou, Emilio. “The Red Flag Act”. The Open University Law School Blog, 2 October 2023. <https://law-school.open.ac.uk/blog/red-flag-act>.

Laird, Elizabeth, Maddy Dwyer, and Hannah Quay-de la Vallee. “Hand in Hand: Schools” Embrace of AI Connected to Increased Risks to Students”. Center for Democracy and Technology, 8 October 2025. <https://cdt.org/insights/hand-in-hand-schools-embrace-of-ai-connected-to-increased-risks-to-students/>.

Lakhani, Nina. “National Science Foundation Staff Decry Trump’s ‘Politically Motivated’ Cuts”. US News. The Guardian, 24 July 2025. <https://www.theguardian.com/us-news/2025/jul/24/national-science-foundation-trumps-cuts>.

Lakoff, George. Don’t Think of an Elephant! Know Your Values and Frame the Debate. Chelsea Green Publishing, 2014. <https://www.chelseagreen.com/product/the-all-new-dont-think-of-an-elephant/>.

Lakoff, George, and Mark Johnson. Metaphors We Live By. University of Chicago Press, 2003. <https://press.uchicago.edu/ucp/books/book/chicago/M/bo3637992.html>.

Landgericht München I. “Urheberrechtsverletzung durch Training einer KI.” Pressemitteilung 11/2025. November 11, 2025. <https://www.justiz.bayern.de/gerichte-und-behoerden/landgericht/muenchen-1/presse/2025/11.php>.

Lawrence, Chan. “Beware General Claims about ‘Generalizable Reasoning Capabilities’ (of Modern AI Systems)”. Alignment Forum, 11 June 2025. <https://www.alignmentforum.org/posts/5uw26uDdFbFQgKzih/beware-general-claims-about-generalizable-reasoning>.

Leake, Matthew. “Are Fears About Online Misinformation in the US Election Overblown? The Evidence Suggests They Might Be”. Reuters Institute for the Study of Journalism, 24 October 2024. <https://reutersinstitute.politics.ox.ac.uk/news/are-fears-about-online-misinformation-us-election-overblown-evidence-suggests-they-might-be>.

Lee, Kai-Fu. “China’s Sputnik Moment and the Sino-American Battle for AI Supremacy”. Asia Society Magazine, 15 March 2022. <https://asiasociety.org/magazine/article/chinas-sputnik-moment-and-sino-american-battle-ai-supremacy>.

Lemley, Mark A., and Bryan Casey. "Fair Learning". SSRN Scholarly Paper No. 3528447. Social Science Research Network, 30 January 2020. <https://doi.org/10.2139/ssrn.3528447>.

Lemoine. "Is LaMDA Sentient? - An Interview". Document Cloud. <https://www.documentcloud.org/documents/22058315-is-lamda-sentient-an-interview/>.

Leppert, Rebecca. "What we know about energy use at U.S. data centers amid the AI boom". Pew Research Center, 24 October 2025. <https://www.pewresearch.org/short-reads/2025/10/24/what-we-know-about-energy-use-at-us-data-centers-amid-the-ai-boom/>.

Levi, Edward H. *An Introduction to Legal Reasoning*. University of Chicago Press, 1949.

Levy, Karen E. C. "The Contexts of Control: Information, Power, and Truck-Driving Work". *The Information Society* 31, no. 2 (2015): 160–74. <https://doi.org/10.1080/01972243.2015.998105>.

Levy, Steven. *In the Plex: How Google Thinks, Works, and Shapes Our Lives*. Simon & Schuster, 2011. https://www.mondothèque.be/wiki/images/2/2b/Levy_Steve_In_The_Plex.pdf.

Lewis, James. "AI and Arms Races". CEPA, 3 June 2025. <https://cepa.org/article/ai-and-arms-races/>.

Lewis, Seth C., David M. Markowitz, and Jon Benedik Bunquin. "Journalists, Emotions, and the Introduction of Generative AI Chatbots: A Large-Scale Analysis of Tweets Before and After the Launch of ChatGPT". arXiv:2409.08761. Preprint, arXiv, 13 September 2024. <https://doi.org/10.48550/arXiv.2409.08761>.

Lichtenberg, Nick. "'It's Almost Tragic': Bubble or Not, the AI Backlash Is Validating One Critic's Warnings". *Fortune*, 24 August 2025. <https://fortune.com/2025/08/24/is-ai-a-bubble-market-crash-gary-marcus-openai-gpt5/>.

LII / Legal Information Institute. "Wire Fraud". https://www.law.cornell.edu/wex/wire_fraud.

Lindsay, James M. "Election 2024: The Deepfake Threat to the 2024 Election". Council on Foreign Relations, 2 February 2024. <https://www.cfr.org/blog/election-2024-deepfake-threat-2024-election>.

Lipsky, Michael. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. Russell Sage Foundation, 1980. <http://archive.org/details/streetlevelburea0000lips>.

Livingstone, Sonia. “iGen: Why Today’s Super-Connected Kids Are Growing up Less Rebellious, More Tolerant, Less Happy – and Completely Unprepared for Adulthood”. *Journal of Children and Media* 12, no. 1 (2018): 118–23. <https://doi.org/10.1080/17482798.2017.1417091>.

LOI N° 2016-1321 Du 7 Octobre 2016 Pour Une République Numérique (1), Pub. L. No. Articles 1 à 39, 2016-1321 (2016). <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746>.

Lomas, Natasha. “ChatGPT Resumes Service in Italy After Adding Privacy Disclosures and Controls”. *TechCrunch*, 28 April 2023. <https://techcrunch.com/2023/04/28/chatgpt-resumes-in-italy/>.

Lombrozo, Tania. “How AI Is Learning to Think on Its Own Like Humans”. *Technology*. *SciTechDaily*, 18 September 2024. <https://scitechdaily.com/how-ai-is-learning-to-think-on-its-own-like-humans/>.

Lowi, Theodore J. “American Business, Public Policy, Case-Studies, and Political Theory”. *World Politics* 16, no. 4 (1964): 677–715. <https://doi.org/10.2307/2009452>.

Lundblad, Nicklas Berild. “Unpredictable Patterns #115: Questioning Our Assumptions”. *Substack newsletter*. *Unpredictable Patterns*, 18 April 2025. <https://unpredictablepatterns.substack.com/p/unpredictable-patterns-115-questioning>.

Luscombe, Richard. “Google Engineer Put on Leave After Saying AI Chatbot Has Become Sentient”. *Technology*. *The Guardian*, 12 June 2022. <https://www.theguardian.com/technology/2022/jun/12/google-engineer-ai-bot-sentient-blake-lemoine>.

Lytton, Charlotte. “AI Hiring Tools May Be Filtering Out the Best Job Applicants”. *BBC*, 16 February 2024. <https://www.bbc.com/worklife/article/20240214-ai-recruiting-hiring-software-bias-discrimination>.

Maas, Matthijs. “AI Is Like... A Literature Review of AI Metaphors and Why They Matter for Policy”. *Institute for Law & AI*, 29 October 2023. <https://law-ai.org/ai-policy-metaphors/>.

Macaulay, Thomas. “Oh Great, the EU Has Ditched Its Facial Recognition Ban”. *The Next Web*, 12 February 2020. <https://thenextweb.com/news/oh-great-the-eu-has-ditched-its-facial-recognition-ban>.

MacGillis, Alec. *Fulfillment: Winning and Losing in One-Click America*. Farrar, Straus and Giroux, 2021. <http://archive.org/details/fulfillmentwinni0000macg>.

- MacKinnon, Rebecca. *Consent of the Networked: The Worldwide Struggle for Internet Freedom*. Basic Books, 2012. <https://consentofthenetworked.com/>.
- Malik, Om. “Decoding Zuck’s Superintelligence Memo”. On My Om, 31 July 2025. <https://om.co/2025/07/30/decoding-zucks-superintelligence-memo/>.
- Mao, Frances. “Robodebt: Illegal Australian welfare hunt drove people to despair”. BBC News, 7 July 2023. <https://www.bbc.com/news/world-australia-66130105>.
- Marcus, Gary. “Deep Learning: A Critical Appraisal”. arXiv:1801.00631. Preprint, arXiv, 2 January 2018. <https://doi.org/10.48550/arXiv.1801.00631>.
- Maxwell, Winston, and Marc Bourreau. “Technology Neutrality in Internet, Telecoms and Data Protection Regulation”. SSRN Scholarly Paper No. 2529680. Social Science Research Network, 23 November 2014. <https://doi.org/10.2139/ssrn.2529680>.
- Maza, Cristina. “Saudi Arabia Gives Citizenship to a Non-Muslim, English-Speaking Robot”. *Newsweek*, 8 November 2017. <https://web.archive.org/web/20171108055113/http://www.newsweek.com/saudi-arabia-robot-sophia-muslim-694152>.
- Mazzucato, Mariana. *The Entrepreneurial State: Debunking Public Vs. Private Sector Myths*. Anthem Press, 2013. <http://archive.org/details/entrepreneurials0000mazz>.
- McAfee, Andrew, and Erik Brynjolfsson. *Machine, Platform, Crowd: Harnessing Our Digital Future*. W. W. Norton & Company, 2017.
- McIntyre, Lee. *Post-Truth*. The MIT Press Essential Knowledge Series. MIT Press, 2018.
- McNamee, Roger. *Zucked; Waking Up To The Facebook Catastrophe*. Penguin Press, 2019. <http://archive.org/details/zucked-waking-up-to-the-facebook-catastrophe-2019-roger-mc-namee>.
- MDR Regulator. “Clinical Trials for Medical Devices & IVD”. Pure Clinical. <https://pureclinical.eu/clinical-trials/>.
- Mehdi, Yusuf. “Announcing Microsoft Copilot, Your Everyday AI Companion”. The Official Microsoft Blog, 21 September 2023. <https://blogs.microsoft.com/blog/2023/09/21/announcing-microsoft-copilot-your-everyday-ai-companion/>.
- Merchant, Brian. “AI Generated Business: The Rise of AGI and the Rush to Find a Working Revenue Model”. AI Now Institute, 21

November 2024. <https://ainowinstitute.org/publications/ai-generated-business>.

Merod, Anna. "Character.AI to Ban Teens from Chatting with Its AI Companions". K-12 Dive, 29 October 2025. <https://www.k12dive.com/news/characterai-to-ban-teens-from-chatting-with-its-ai-companions/804199/>.

Messier, Ashlyn. "What Is Google Bard? How the AI Chatbot Works, How to Use It and Why It's Controversial". Fox News, Fox News, 21 April 2023. <https://www.foxnews.com/tech/what-google-bard-how-the-ai-chatbot-works-how-use-it-why-its-controversial>.

Mills, Anna, and Nate Angell. "Are We Tripping? The Mirage of AI Hallucinations". SSRN Scholarly Paper No. 5127162. Social Science Research Network, 6 February 2025. <https://doi.org/10.2139/ssrn.5127162>.

Ministry of Foreign Affairs of the People's Republic of China. "Global AI Governance Action Plan". 26 July 2025. https://www.mfa.gov.cn/eng/xw/zyxw/202507/t20250729_11679232.html.

Mitchell, Melanie. "The Metaphors of Artificial Intelligence". *Science* 386, no. 6723 (2024): eadt6140. <https://doi.org/10.1126/science.adt6140>.

Mitchell, Melanie. "Why AI Is Harder Than We Think". arXiv:2104.12871. Preprint, arXiv, 28 April 2021. <https://doi.org/10.48550/arXiv.2104.12871>.

Mokyr, Joel, Chris Vickers, and Nicolas L. Ziebarth. "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?" *Journal of Economic Perspectives* 29, no. 3 (2015): 31–50. <https://doi.org/10.1257/jep.29.3.31>.

Mollman, Steve. "OpenAI CEO Sam Altman: A.I. 'Good at Doing Tasks' but Terrible at Doing Whole Jobs—for Now". *Fortune*, 8 June 2023. <https://fortune.com/2023/06/08/openai-ceo-sam-altman-ai-good-at-tasks-bad-at-jobs-for-now/>.

"Montréal Declaration for a Responsible Development of Artificial Intelligence, Launched November 3, 2017; Official Release December 4, 2018". *Déclaration de Montréal IA Responsable*, n.d. <https://montrealdeclaration-responsibleai.com/the-declaration/>.

Moody, Glyn. "European Publishers Council Stays True to the Tired Old Trope about 'Copyright Theft'". *Walled Culture*, 28 May 2025. <https://walledculture.org/european-publishers-council-stays-true-to-the-tired-old-trope-about-copyright-theft>.

- Morozov, Evgeny. To Save Everything, Click Here: The Folly of Technological Solutionism. PublicAffairs, 2013. <http://archive.org/details/tosaveeverything0000moro>.
- Morris, Meredith Ringel, Jascha Sohl-Dickstein, Noah Fiedel, et al. “Levels of AGI for Operationalizing Progress on the Path to AGI”. Version 5. Preprint, arXiv, 2023. <https://doi.org/10.48550/ARXIV.2311.02462>.
- Morrison, Ed. “Complexity Thinking. What Is It?” Agile Strategy Lab, 30 August 2023. <https://agilestrategylab.org/complexity-thinking-what-is-it/>.
- Morriss, Andrew P. “The Wild West Meets Cyberspace”. Foundation for Economic Education, 7 January 1998. <https://fee.org/articles/the-wild-west-meets-cyberspace/>.
- Mulligan, Deirdre K., Joshua A. Kroll, Nitin Kohli, and Richmond Y. Wong. “This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology”. *Proceedings of the ACM on Human-Computer Interaction* 3, no. CSCW (2019): 1–36. <https://doi.org/10.1145/3359221>.
- Municipality of Amsterdam. Grip on Algorithms: Approach and Tools for a Responsible Use of Algorithms in Amsterdam. Amsterdam: City of Amsterdam, 2022. https://www.amsterdam.nl/publish/pages/1053010/playbook_algorithms.pdf.
- Murgia, Madhumita. “Sci-Fi Writer Ted Chiang: ‘The Machines We Have Now Are Not Conscious’”. *Financial Times*, 2 June 2023. <https://www.ft.com/content/c1f6d948-3dde-405f-924c-09cc0dcf8c84>.
- Murgia, Madhumita, Cristina Criddle, and Melissa Heikkilä. “AI pioneers claim human-level general intelligence is already here.” *Financial Times*. 6 November 2025. <https://www.ft.com/content/5f2f411c-3600-483b-bee8-4f06473ecdc0>.
- Musk, Elon and Neuralink. “An Integrated Brain-Machine Interface Platform With Thousands of Channels”. *Journal of Medical Internet Research* 21, no. 10 (2019): e16194. <https://doi.org/10.2196/16194>.
- Musto, Julia. “Three Mile Island nuclear plant, site of worst US reactor accident, to reopen for Microsoft”. *The Independent*, 20 September 2024. <https://www.independent.co.uk/climate-change/microsoft-three-mile-island-nuclear-pennsylvania-b2616314.html>.
- Nakamoto, Satoshi. “Bitcoin: A Peer-to-Peer Electronic Cash System”. Nakamoto Institute, 31 October 2008. <https://nakamotoinstitute.org/library/bitcoin/>.

Narayanan, Arvind, and Sayash Kapoor. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton University Press, 2024. <https://press.princeton.edu/books/hardcover/9780691249131/ai-snake-oil>.

National Academies of Sciences, Engineering, and Medicine. *Securing the Vote: Protecting American Democracy*. National Academies Press, 2018. <https://doi.org/10.17226/25120>.

National Academy of Sciences. *At the Nexus of Cybersecurity and Public Policy: Some Basic Concepts and Issues*. National Academies Press, 2014. <https://doi.org/10.17226/18749>.

National Science & Technology Council. *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. Select Committee on AI, 2019. <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>.

National Security Commission on Artificial Intelligence. "Final Report". March 2021. <https://reports.nscai.gov/final-report/>.

National Security Commission on Artificial Intelligence (U.S.). "Final Report: National Security Commission on Artificial Intelligence". Report. UNT Digital Library, National Security Commission on Artificial Intelligence (U.S.), 1 March 2021. United States. <https://digital.library.unt.edu/ark:/67531/metadc1851188/>.

Naughton, Barry J. *The Chinese Economy: Adaptation and Growth*. 2nd edn. MIT Press, 2016.

Nerlich, Brigitte. "Hunting for AI Metaphors". *Making Science Public*, 12 April 2024. <https://blogs.nottingham.ac.uk/makingsciencepublic/2024/04/12/hunting-for-ai-metaphors/>.

Newsdesk. "UAE Explores AI Judges". *Gulf Insight 360*, 7 February 2025. <https://gulfsight360.com/uae-explores-ai-judges-a-bold-step-towards-judicial-innovation/?amp=1>.

"New York State Assembly Bill A6453A, Relates to the Training and Use of Artificial Intelligence Frontier Models, 2025-2026 Leg". <https://www.nysenate.gov/legislation/bills/2025/A6453/amendment/A>.

Nissenbaum, H. "How Computer Systems Embody Values". *Computer* 34, no. 3 (2001): 120–119. <https://doi.org/10.1109/2.910905>.

Nix, Naomi, Cat Zakrzewski, and Gerrit De Vynck. "Silicon Valley Is Pricing Academics out of AI Research". *The Washington Post*, 10 March 2024. <https://www.washingtonpost.com/technology/2024/03/10/big-tech-companies-ai-research/>.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018. <https://doi.org/10.18574/nyu/9781479833641.001.0001>.

Noveck, Beth Simone. "Crowdlaw: Collective Intelligence and Lawmaking". *Analyse & Kritik* 40, no. 2 (2018): 359–80. <https://doi.org/10.1515/auk-2018-0020>.

Novelli, Claudio, Luciano Floridi, and Giovanni Sartor. *AI as Legal Persons: Past, Patterns, and Prospects*. n.d. <https://a-mcc.eu/wp-content/uploads/2024/11/ssrn-5032265.pdf>.

Oduro, Serena, Emanuel Moss, and Jacob Metcalf. "Obligations to Assess: Recent Trends in AI Accountability Regulations". *Patterns* 3, no. 11 (2022): 100608. <https://doi.org/10.1016/j.patter.2022.100608>.

OECD. "Case Studies on the Regulatory Challenges Raised by Innovation and the Regulatory Responses". OECD, 13 December 2021. https://www.oecd.org/en/publications/case-studies-on-the-regulatory-challenges-raised-by-innovation-and-the-regulatory-responses_8fa190b5-en.html.

Ojha, Saanya. "The 95% Problem: AI Isn't Overhyped, Enterprises Are Underprepared". Substack newsletter. *The Change Constant*, 21 August 2025. <https://saanyaojha.substack.com/p/the-95-problem-ai-isnt-overhyped>.

Olah, Chris, Nick Cammarata, Ludwig Schubert, Gabriel Goh, Michael Petrov, and Shan Carter. "Zoom In: An Introduction to Circuits". *Distill* 5, no. 3 (2020): e00024.001. <https://doi.org/10.23915/distill.00024.001>.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, 2016.

Onslow, Mike. "The Phantom Menace: AI Hallucinations and Their Impact". *Artificial Antics*, 12 March 2024. <https://antics.tv/pi/understanding-ai-hallucination-insights-and-implications>.

OpenAI. "About". 30 September 2025. <https://openai.com/about/>.

OpenAI. "Introducing ChatGPT". 30 November 2022. <https://openai.com/index/chatgpt/>.

OpenAI. "Key Concepts". OpenAI API. <https://platform.openai.com>.

OpenAI. "Preparedness Framework". 29 September 2025. <https://openai.com/safety/>.

Ord, Toby. *The Precipice: Existential Risk and the Future of Humanity*. Hachette Books, 2020. <https://www.hachettebookgroup.com/titles/toby-ord/the-precipice/9780316484893/?lens=grand-central-publishing>.

Ortutay, Barbara. “What You Should Know About Section 230, the Rule That Shaped Today’s Internet”. PBS News, 21 February 2023. <https://www.pbs.org/newshour/politics/what-you-should-know-about-section-230-the-rule-that-shaped-todays-internet>.

Osenga, Kristen. “The Internet Is Not a Super Highway: Using Metaphors to Communicate Information and Communications Policy”. *Journal of Information Policy* 3 (January 2013): 30–54. <https://doi.org/10.5325/jinfopoli.3.2013.0030>.

Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

O’Sullivan, Isobel. “What Is Dead Internet Theory and Could It Change the Web?” *Tech.Co*, 22 May 2024. <https://tech.co/news/what-is-dead-internet-theory>.

Palmer, Shelly. “America’s AI Action Plan: What You Need to Know”. Shelly Palmer, 24 July 2025. <https://shellypalmer.com/2025/07/americas-ai-action-plan-what-you-need-to-know/>.

Panetta, Leon E. *Defending the Nation from Cyber Attack* (Business Executives for National Security). U.S. Department of Defense, 2012. <https://nsarchive.gwu.edu/sites/default/files/documents/2700165/Document-78.pdf>.

Park, Sumner, and Lydia Hu. “School Districts Blocking ChatGPT Amid Fears of Cheating, Educators Weigh in on AI”. *Text.Article*. FOXBusiness, Fox Business, 11 January 2023. <https://www.foxbusiness.com/technology/school-districts-blocking-chatgpt-amid-fears-of-cheating-educators-weigh-in-on-ai>.

Park, Suzy E. *Technological Convergence: Regulatory, Digital Privacy, and Data Security Issues*. CRS Report R45746. Congressional Research Service, 2019. https://www.everycrsreport.com/files/20190530_R45746_461fe45bf0ce4a3f25e8d1a507c308fee4d5bfd4.pdf.

Participedia. “Participatory Consensus Conferences”. Participedia. <https://participedia.net/method/participatory-consensus-conferences>.

Partridge, Joanna. “Global stock markets fall sharply over AI bubble fears”. *The Guardian*, 5 November 2025. <https://www.theguardian>.

com/business/2025/nov/05/global-stock-markets-fall-sharply-over-ai-bubble-fears.

Pasquale, Frank. "The Algorithmic Self". *The Hedgehog Review*, 2015. <https://hedgehogreview.com/issues/too-much-information/articles/the-algorithmic-self>.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.

Pearlman, Jonathan. "Australian States Block ChatGPT in Schools Even as Critics Say Ban Is Futile". *The Straits Times*, 6 January 2023. <https://www.straitstimes.com/asia/australianz/australian-states-block-chatgpt-in-schools-even-as-critics-say-ban-is-futile>.

Pelopidas, Benoît. "The Oracles of Proliferation: How Experts Maintain a Biased Historical Reading That Limits Policy Innovation". *The Nonproliferation Review* 18, no. 1 (2011): 297–314. <https://doi.org/10.1080/10736700.2011.549185>.

Perez, Sarah. "Character.AI, the A16z-Backed Chatbot Startup, Tops 1.7M Installs in First Week". *TechCrunch*, 31 May 2023. <https://techcrunch.com/2023/05/31/character-ai-the-a16z-backed-chatbot-startup-tops-1-7m-installs-in-first-week/>.

Pierson, Paul. "Increasing Returns, Path Dependence, and the Study of Politics". *The American Political Science Review* 94, no. 2 (2000): 251–67. <https://doi.org/10.2307/2586011>.

Placani, Adriana. "Anthropomorphism in AI: Hype and Fallacy". *AI and Ethics* 4, no. 3 (2024): 691–98. <https://doi.org/10.1007/s43681-024-00419-4>.

Placido, Dani Di. "Facebook's AI-Generated 'Shrimp Jesus,' Explained". *Forbes*, 28 April 2024. <https://www.forbes.com/sites/danidiplacido/2024/04/28/facebooks-surreal-shrimp-jesus-trend-explained/>.

Plato. *Phaedrus*. Translated by Benjamin Jowett. Project Gutenberg, 2016. <https://www.gutenberg.org/files/1636/1636-h/1636-h.htm>.

"Playing to the Crowd: Musicians, Audiences, and the Intimate Work of Connection". NYU Press, n.d. <https://nyupress.org/9781479821587/playing-to-the-crowd/>.

Polanyi, Michael. *The Tacit Dimension*. [1st ed.]. Doubleday, 1966. https://openlibrary.org/books/OL5990259M/The_tacit_dimension.

Postman, Neil. *Technopoly: The Surrender of Culture to Technology*. Vintage Books, 1993. <https://ia601802.us.archive.org/27/items/235p-technopoly-neil-postman/235p%20technopoly-neil-postman.pdf>.

Pratley, Nils. “The AI Valuation Bubble Is Now Getting Silly”. *Business*. The Guardian, 8 October 2025. <https://www.theguardian.com/technology/nils-pratley-on-finance/2025/oct/08/the-ai-valuation-bubble-is-now-getting-silly>.

Predpol. “Predictive Policing”. Upturn. <https://teamupturn.gitbooks.io/predictive-policing/content/systems/predpol.html>.

Purnell, Newley, and Parmy Olson. “AI Startup Boom Raises Questions of Exaggerated Tech Savvy”. *Tech*. Wall Street Journal, 14 August 2019. <https://www.wsj.com/articles/ai-startup-boom-raises-questions-of-exaggerated-tech-savvy-11565775004>.

Putnam, Robert D. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, 2000. <https://www.simonandschuster.com/books/Bowling-Alone-Revised-and-Updated/Robert-D-Putnam/9781982130848>.

Qiao, Tingting, Jing Zhang, Duanqing Xu, and Dacheng Tao. “Learn, Imagine and Create: Text-to-Image Generation from Prior Knowledge”. In *Advances in Neural Information Processing Systems*, edited by H. Wallach, H Larochelle, A Beygelzimer, F d’Alché-Buc, E Fox, and R Garnett, vol. 32. Curran Associates, Inc., 2019. https://papers.nips.cc/paper_files/paper/2019/hash/d18f655c3fce66ca401d5f38b48c89af-Abstract.html.

Rasenberger, Will. “Why Metaphors Matter in AI Law and Policy”. *The Regulatory Review*, 2 May 2024. <https://www.theregreview.org/2024/05/02/rasenberger-why-metaphors-matter-in-ai-law-and-policy/>.

Rbanffy. “Crypto’s Wild West Era Is Over”. *Hacker News*, n.d. <https://news.ycombinator.com/item?id=44603096>.

Regulation (EU) 2024/1689. “Article 13: Transparency and Provision of Information to Deployers”. *EU Artificial Intelligence Act*, 2 February 2025. <https://artificialintelligenceact.eu/article/13/>.

Renzella, Jake, and Vlada Rozova. “The ‘Dead Internet Theory’ Makes Eerie Claims About an AI-Run Web. The Truth Is More Sinister”. *The Conversation*, 20 May 2025. <https://doi.org/10.64628/AA.n5axxgun6>.

Renzi, Barbara Gabriella, Matthew Cotton, Giulio Napolitano, and Ralf Barkemeyer. “Rebirth, Devastation and Sickness: Analyzing

the Role of Metaphor in Media Discourses of Nuclear Power”. *Environmental Communication* 11, no. 5 (2017): 624–40. <https://doi.org/10.1080/17524032.2016.1157506>.

Reuters. “Most coal-fired power plants will delay retirement to feed AI boom, energy secretary says”. *Mining Weekly*, 26 September 2025. <https://www.miningweekly.com/article/most-coal-fired-power-plants-will-delay-retirement-to-feed-ai-boom-energy-secretary-says-2025-09-26>.

Rich, Andrew. *Think Tanks, Public Policy, and the Politics of Expertise*. Cambridge University Press, 2004. <http://archive.org/details/thinktankspublic00andr>.

Robbins, Jacob. “AI Startups Grabbed a Third of Global VC Dollars in 2024”. *PitchBook*, 9 January 2025. <https://pitchbook.com/news/articles/ai-startups-grabbed-a-third-of-global-vc-dollars-in-2024>.

Roberts, Sarah T. *Behind the Screen: Content Moderation in the Shadows of Social Media*. Yale University Press, 2019. <https://doi.org/10.2307/j.ctvhrcz0v>.

Robertson, David. “Ten Reasons to Say No: A Primer on Sidewalk Labs’ Plan for Toronto”. *Socialist Project*, 14 October 2019. <https://socialistproject.ca/2019/10/ten-reasons-to-say-no-sidewalk-lab-toronto/>.

“Robots: Legal Affairs Committee Calls for EU-Wide Rules | News | European Parliament”. 12 January 2017. <https://www.europarl.europa.eu/news/en/press-room/20170110IPR57613/robots-legal-affairs-committee-calls-for-eu-wide-rules>.

Rodrigo, Chris Mills. “How Major Labor Unions Are Positioning on AI”. *Tech Policy Press*, 28 October 2025. <https://techpolicy.press/how-major-labor-unions-are-positioning-on-ai>.

Roose, Kevin. “An A.I.-Generated Picture Won an Art Prize. Artists Aren’t Happy”. *Technology*. *The New York Times*, 2 September 2022. <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>.

Roose, Kevin. “If A.I. Systems Become Conscious, Should They Have Rights?” *Technology*. *The New York Times*, 24 April 2025. <https://www.nytimes.com/2025/04/24/technology/ai-welfare-anthropic-claude.html>.

Rosen, Zvi S. “The Unoriginal Sin of Technology-Neutral Copyright”. *Minnesota Law Review* 100, no. 1 (2016): 1495–558.

Ross, Casey, and Ike Swetlitz. "IBM Pitched Its Watson Supercomputer as a Revolution in Cancer Care. It's Nowhere Close". STAT, 5 September 2017. <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>.

Rossetti, Philip. "What Happened to Electricity Becoming 'Too Cheap to Meter?'" American Action Forum, 30 August 2017. <https://www.americanactionforum.org/daily-dish/happened-electricity-becoming-cheap-meter/>.

Roush, Tyler. "Nvidia Becomes First Company Worth \$5 Trillion". Forbes, 29 October 2025. <https://www.forbes.com/sites/tylerroush/2025/10/29/nvidia-becomes-first-company-worth-5-trillion/>.

Russell, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking Press, 2019. <https://www.penguinrandomhouse.com/books/566677/human-compatible-by-stuart-russell/>.

Sætra, Henrik Skaug, and Evan Selinger. "Technological Remedies for Social Problems: Defining and Demarcating Techno-Fixes and Techno-Solutionism". *Science and Engineering Ethics* 30, no. 6 (2024): 60. <https://doi.org/10.1007/s11948-024-00524-x>.

Salvaggio, Eryk. "The Black Box Myth: What the Industry Pretends Not to Know About AI". Tech Policy Press, 17 June 2025. <https://techpolicy.press/the-black-box-myth-what-the-industry-pretends-not-to-know-about-ai>.

Samuel, Sigal. "'AI Will Kill Everyone' Is Not an Argument. It's a Worldview". Vox, 17 September 2025. <https://www.vox.com/future-perfect/461680/if-anyone-builds-it-yudkowsky-soares-ai-risk>.

Samuelson, Pamela. "Does Using In-Copyright Works as Training Data Infringe?" Wolters Kluwer Copyright Blog, 18 July 2025. <https://legalblogs.wolterskluwer.com/copyright-blog/does-using-in-copyright-works-as-training-data-infringe/>.

Scharre, Paul. *Four Battlegrounds: Power in the Age of Artificial Intelligence*. W. W. Norton & Company, 2023.

Schmidt, Eric, dir. *The AI Revolution Is Underhyped*. TED Talk, 2025. Video, 1537. https://www.ted.com/talks/eric_schmidt_the_ai_revolution_is_underhyped.

Schmidt, Eric, and Selina Xu. "Silicon Valley Is Drifting Out of Touch With the Rest of America". Opinion. *The New York Times*, 19 August 2025. <https://www.nytimes.com/2025/08/19/opinion/artificial-general-intelligence-superintelligence.html>.

Schneid, Rebecca, and Andrew R. Chow. "How Trump's Use of AI Videos Is Changing His Political Playbook". *TIME*, 21 October 2025. <https://time.com/7327317/trump-ai-video-political-weapon/>.

Schneier, Bruce. *Click Here to Kill Everybody: Security and Survival in a Hyperconnected World*. W. W. Norton & Company, 2018. <https://www.schneier.com/books/click-here/>.

Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. W. W. Norton & Company, 2015. <https://www.schneier.com/books/data-and-goliath/>.

Scholz, Trebor. *Platform Cooperativism: Challenging the Corporate Sharing Economy*. Rosa Luxemburg Stiftung, 2016. https://rosalux.nyc/wp-content/uploads/2020/11/RLS-NYC_platformcoop.pdf.

Schüll, Natasha Dow. "Data for Life: Wearable Technology and the Design of Self-Care". *BioSocieties* 11, no. 3 (2016): 317–33. <https://doi.org/10.1057/biosoc.2015.47>.

Schwab, Klaus. *The Fourth Industrial Revolution*. World Economic Forum, 2016. https://law.unimelb.edu.au/__data/assets/pdf_file/0005/3385454/Schwab-The_Fourth_Industrial_Revolution_Klaus_S.pdf.

Selinger, Evan. "The Delusion at the Center of the A.I. Boom". *Slate*, 29 March 2023. <https://slate.com/technology/2023/03/chatgpt-artificial-intelligence-solutionism-hype.html>.

Selwyn, Neil. "Lessons to Be Learnt? Education, Techno-Solutionism, and Sustainable Development". In *Technology and Sustainable Development*, 1st edn, by Henrik Skaug Sætra. Routledge, 2023. <https://doi.org/10.1201/9781003325086-6>.

Shapin, Steven, and Simon Schaffer. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press, 1985.

Shapiro, Gary. *The Comeback: How Innovation Will Restore the American Dream*. Beaufort Books, 2011. <http://archive.org/details/comeback0000shap>.

Shardlow, Matthew, and Piotr Przybyła. "Deanthropomorphising NLP: Can a Language Model Be Conscious?" *PLOS ONE* 19, no. 12 (2024): e0307521. <https://doi.org/10.1371/journal.pone.0307521>.

Sharkey, Catherine. "Products Liability for Artificial Intelligence". *Lawfare*, Lawfare, 25 September 2024. <https://www.lawfaremedia.org/article/products-liability-for-artificial-intelligence>.

Sharwood, Simon. "AWS CEO Says AI Replacing Junior Staff Is 'Dumbest Idea'". *The Register*, 21 August 2025. https://www.theregister.com/2025/08/21/aws_ceo_entry_level_jobs_opinion/.

Shepardson, David. "US Finalizes \$6.6 Billion Chips Award for TSMC Ahead of Trump Return". *Technology*. Reuters, 15 November 2024. <https://www.reuters.com/technology/us-finalizes-66-billion-chips-award-tsmc-ahead-trump-return-2024-11-15/>.

Shepherd, Brittany, and Monica Madden. "Gavin Newsom Is Gambling on New Rules: 'Trump 2.0'". *ABC News*, 2 November 2025. <https://abcnews.go.com/US/gavin-newsom-gambling-new-rules-trump-20/story?id=126900405>.

Signé, Landry, and Colette van der Ven. "Turning Policy into Action: Overcoming Policy Failures and Bridging Implementation Gaps". *Brookings*, 28 February 2024. <https://www.brookings.edu/articles/turning-policy-into-action-in-africa/>.

Singapore Government, Smart Nation and Digital Government Office. "National AI Strategy". *Smart Nation Singapore*, 7 October 2025. <https://www.smartnation.gov.sg/initiatives/national-ai-strategy/>.

Singer, Peter W., and Allan Friedman. *Cybersecurity and Cyberwar: What Everyone Needs to Know®*. What Everyone Needs To Know®. Oxford University Press, 2014.

Singla, Alex, Alexander Sukharevsky, Lareina Yee, and Michael Chui, with Bryce Hall and Tara Balakrishnan. "The state of AI in 2025: Agents, innovation, and transformation." *McKinsey & Company*. 5 November 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>.

Skeem, Jennifer L., and Christopher T. Lowenkamp. "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact". *Criminology* 54, no. 4 (2016): 680–712. <https://doi.org/10.1111/1745-9125.12123>.

Skocpol, Theda. *Diminished Democracy: From Membership to Management in American Civic Life*. Norman : University of Oklahoma Press, 2003. <http://archive.org/details/diminisheddemocr0000skoc>.

Smith, Noah. "America's Future Could Hinge on Whether AI Slightly Disappoints". *Noahpinion*, 12 October 2025. <https://www.noahpinion.blog/p/americas-future-could-hinge-on-whether>.

Sontag, Susan. *On Photography*. Farrar, Straus and Giroux, 1977.

Soper, Taylor. "Seattle Public Schools Bans ChatGPT; District 'Requires Original Thought and Work from Students'". *GeekWire*,

16 January 2023. <https://www.geekwire.com/2023/seattle-public-schools-bans-chatgpt-district-requires-original-thought-and-work-from-students/>.

Soydaner, Derya. "Attention Mechanism in Neural Networks: Where It Comes and Where It Goes". *Neural Computing and Applications* 34, no. 16 (2022): 13371–85. <https://doi.org/10.1007/s00521-022-07366-3>.

Srebrnik, Dana. "The Dead Internet Theory and the Silent Surge of Bots". Wix Blog, 18 September 2024. <https://www.wix.com/blog/dead-internet-theory>.

St. Jacques, Michael. "Their Swords, Our Plowshares: 'Peaceful' Nuclear Weapons, Propaganda, and Cold War Memory Expressed in Film: 1959-1989". Master's Thesis, James Madison University, 2016. <https://commons.lib.jmu.edu/master201019/102>.

Steering Group of the Artificial Intelligence Programme. "Finland's Age of Artificial Intelligence: Turning Finland into a Leading Country in the Application of Artificial Intelligence". Publications of the Ministry of Economic Affairs and Employment 47/2017, October 2017. https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf.

Stone, Brad. *The Upstarts: How Uber, Airbnb, and the Killer Companies of the New Silicon Valley Are Changing the World*. Little, Brown and Company, 2017. <http://archive.org/details/upstartshowubera0000ston>.

Stone, Deborah. *Policy Paradox: The Art of Political Decision Making*. W.W. Norton & Company, 2012. <http://archive.org/details/policyparadoxart0000ston>.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP". *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2019, 3645–50. <https://doi.org/10.18653/v1/P19-1355>.

Suleyman, Mustafa. "We Must Build AI for People; Not to Be a Person". Mustafa Suleyman AI, 19 August 2025. <https://mustafa-suleyman.ai/seemingly-conscious-ai-is-coming>.

Syed, Tabrez. "The AI Gold Rush: Are We Mining or Just Selling Shovels?" BoxCars AI, 4 April 2024. <https://blog.boxcars.ai/p/the-ai-gold-rush-are-we-mining-or>.

Texas Legislature. "Texas Legislature, Senate Bill No. 751 (86th Leg., Regular Session): Relating to the Creation of a Criminal Offense for Fabricating a Deceptive Video with Intent to Influence the Outcome of an Election, Enacted June 14, 2019". 1 September 2019. <https://capitol.texas.gov/BillLookup/History.aspx?LegSess=86R&Bill=SB751>.

The Copyright Act of 1909 (as Amended and Codified), Title 17 U.S.C. n.d. https://ipmall.law.unh.edu/sites/default/files/hosted_resources/lipa/copyrights/CopyrightAct1909amendedcodified_.pdf.

The Editorial Board. "AI Should Not Be a Black Box". Financial Times, 30 May 2024. <https://www.ft.com/content/9378339f-a0aa-434a-a687-5dd9a13df5fe>.

The Nobel Prize. "Nobel Prize in Chemistry 2024". 2024. <https://www.nobelprize.org/prizes/chemistry/2024/summary/>.

Thuronyi, George. "Copyright Law and New Technologies: A Long and Complex Relationship". Webpage. The Library of Congress, 22 May 2017. <https://blogs.loc.gov/copyright/2017/05/copyright-law-and-new-technologies-a-long-and-complex-relationship>.

Tiku, Nitasha. "The Google Engineer Who Thinks the Company's AI Has Come to Life". The Washington Post, 11 June 2022. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>.

Toews, Rob. "Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared". Forbes, 25 May 2020. <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/>.

Topol, Eric. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019.

Torres, Emile P. "The Dangerous Ideas of 'Longtermism' and 'Existential Risk'". Current Affairs, 28 July 2021. <https://www.currentaffairs.org/news/2021/07/the-dangerous-ideas-of-longtermism-and-existential-risk>.

Tracy, James F. "Review: C. Edwin Baker, *Media Concentration and Democracy: Why Ownership Matters*. New York: Cambridge University Press, 2007. £35.00 (Hbk), £14.99 (Pbk). 256 Pp". *European Journal of Communication* 23, no. 2 (2008): 249–53. <https://doi.org/10.1177/02673231080230020610>.

Tuquero, Loreben. "Trump's White House Is Testing the Limits of AI-Driven Political Messaging". Poynter, 28 October 2025. <https://>

www.poynter.org/fact-checking/2025/trump-white-house-ai-political-messaging/.

Turnitin. "Turnitin Turns on AI Writing Detection Capabilities for Educators and Institutions (NA and EMEA)". 4 April 2023. <https://www.turnitin.co.uk/press/turnitin-turns-on-ai-writing-detection-capabilities-for-educators-and-institutions>.

Tversky, Amos, and Daniel Kahneman. "Availability: A Heuristic for Judging Frequency and Probability". *Cognitive Psychology* 5, no. 2 (1973): 207–32. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9).

U.K. Financial Conduct Authority. "Regulatory Sandbox". FCA, 1 March 2022. <https://www.fca.org.uk/firms/innovation/regulatory-sandbox>.

UK Government - Department for Science, Innovation & Technology and Office for Artificial Intelligence. "AI Regulation: A Pro-Innovation Approach". Government of UK, 3 August 2023. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>.

Uniform Law Commission. "Uniform Commercial Code". <https://www.uniformlaws.org/acts/ucc>.

United Nations Environment Programme. "The Montreal Protocol on Substances That Deplete the Ozone Layer". Ozone Secretariat. <https://ozone.unep.org/treaties/montreal-protocol>.

United States Congress. "Federal Criminal Code, Title 18, Part I, Chapter 63 & Chapter 113". https://www.oas.org/juridico/spanish/cyber/us_cyb_law_fed_crim_code.pdf.

U.S. Congress, Office of Technology Assessment. "The OTA Legacy". Princeton University. <http://www.princeton.edu/~ota/>.

U.S. Congressional Budget Office. "Introduction to CBO". <https://www.cbo.gov/about/overview>.

U.S. Food & Drug Administration. "Breakthrough Devices Program". FDA, FDA, 20 August 2025. <https://www.fda.gov/medical-devices/how-study-and-market-your-device/breakthrough-devices-program>.

U.S. Food & Drug Administration. "Classify Your Medical Device". FDA, FDA, 14 August 2020. <https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device>.

U.S. Food & Drug Administration. "The Drug Development Process – Step 3: Clinical Research". FDA, FDA, 18 April 2019. <https://www.fda.gov/patients/drug-development-process/step-3-clinical-research>.

U.S. House Committee on Oversight and Accountability. Oversight of AI: Rules for Artificial Intelligence” (Hearing before the Subcommittee on Privacy, Technology, and the Law, Senate Judiciary Committee, May 16, 2023). Gibson Dunn Public Policy Alert, 2023. <https://www.gibsondunn.com/wp-content/uploads/2023/06/oversight-of-ai-rules-for-artificial-intelligence-and-artificial-intelligence-in-government-hearings.pdf>.

U.S. National Highway Traffic Safety Administration. “Laws & Regulations”. Text. NHTSA. <https://www.nhtsa.gov/laws-regulations>.

U.S. Senate. Committee on Commerce, Science, and Transportation. “Video Game Violence”. C-SPAN, 28 October 2025. <https://www.c-span.org/program/senate-committee-video-game-violence/129470>.

U.S. Food & Drug Administration. “Safety Management System (SMS)”. Federal Aviation Administration. <https://www.faa.gov/about/initiatives/sms>.

Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press, 2016.

Van Der Velden, Maja, and Christina Mörtberg. “Participatory Design and Design for Values”. In *Handbook of Ethics, Values, and Technological Design*, edited by Jeroen Van Den Hoven, Pieter E. Vermaas, and Ibo Van De Poel. Springer Netherlands, 2015. https://doi.org/10.1007/978-94-007-6970-0_33.

Verbeek, Peter-Paul, and Robert P. Crease. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. Penn State University Press, 2005. <https://www.jstor.org/stable/10.5325/j.ctv14gp4w7>.

Video Viory. “Congratulations! - EC Spox on “Pregnancy” of Albanian AI Minister, Dodges Questions on US-China Trade Talks”. 30 October 2025. https://www.viory.video/en/videos/a3454_30102025/congratulations-ec-spox-on-pregnancy-of-albanian-ai-minister-dodges-questions-on-us-china-trade-talks.

Virki, Tarmo. “Finland Seeks to Teach 1% of All Europeans Basics on AI”. *World*. Reuters, 3 January 2020. <https://www.reuters.com/article/world/finland-seeks-to-teach-1-of-all-europeans-basics-on-ai-idUSKBN1YE1CT/>.

von der Leyen, Ursula. *A Union That Strives for More. My Agenda for Europe - Political Guidelines for the Next Commission*. European Commission, n.d. <https://commission.europa.eu/>

document/download/063d44e9-04ed-4033-acf9-639ecb187e87_en?filename=political-guidelines-next-commission_en.pdf.

von der Leyen, Ursula. "Speech by President von der Leyen at the Annual EU Budget Conference 2025." European Commission. 19 May 2025. https://ec.europa.eu/commission/presscorner/detail/en/speech_25_1284.

vTaiwan. "vTaiwan: rethinking democracy". <https://info.vtaiwan.tw/>.

Walker, Brian, and David Salt. *Resilience Thinking: Sustaining Ecosystems and People in a Changing World*. Island Press, 2006.

Wang, Maya. "China's Chilling 'Social Credit' Blacklist". Opinion. Wall Street Journal, 11 December 2017. <https://www.wsj.com/articles/chinas-chilling-social-credit-blacklist-1513036054>.

Watters, Audrey. "A Brief History of Calculators in the Classroom". Hack Education, 12 March 2015. <http://hackededucation.com/2015/03/12/calculators>.

Weil, Jonathan. "Is the Flurry of Circular AI Deals a Win-Win—or Sign of a Bubble?". The Wall Street Journal, 22 October 2025. <https://www.wsj.com/tech/ai/is-the-flurry-of-circular-ai-deals-a-win-win-or-sign-of-a-bubble-8a2d70c5>.

Weiss-Blatt, Nirit. "Why The 'Doom Bible' Left Many Reviewers Unconvinced". AI Panic, 1 November 2025. <https://www.aipanic.news/p/why-the-doom-bible-left-many-reviewers>.

Wikipedia. "AI Winter". https://en.wikipedia.org/w/index.php?title=AI_winter&oldid=1309819497.

Wikipedia. "Artificial General Intelligence". https://en.wikipedia.org/w/index.php?title=Artificial_general_intelligence&oldid=1320111828.

Wikipedia. "Artificial Intelligence Act". https://en.wikipedia.org/w/index.php?title=Artificial_Intelligence_Act&oldid=1318974732.

Wikipedia. "Availability Heuristic". https://en.wikipedia.org/w/index.php?title=Availability_heuristic&oldid=1271977939.

Wikipedia. "Backpropagation". <https://en.wikipedia.org/w/index.php?title=Backpropagation&oldid=1314002217>.

Wikipedia. "Chief Digital and Artificial Intelligence Office". https://en.wikipedia.org/w/index.php?title=Chief_Digital_and_Artificial_Intelligence_Office&oldid=1304971117.

Wikipedia. “Churnalism”. <https://en.wikipedia.org/w/index.php?title=Churnalism&oldid=1320447169>.

Wikipedia. “Clarke’s Three Laws”. https://en.wikipedia.org/w/index.php?title=Clarke%27s_three_laws&oldid=1307803560.

Wikipedia. “COMPAS (software)”. [https://en.wikipedia.org/w/index.php?title=COMPAS_\(software\)&oldid=1314390733](https://en.wikipedia.org/w/index.php?title=COMPAS_(software)&oldid=1314390733).

Wikipedia. “Dead Internet Theory”. https://en.wikipedia.org/w/index.php?title=Dead_Internet_theory&oldid=1321027886.

Wikipedia. “Dutch childcare benefits scandal”. https://en.wikipedia.org/wiki/Dutch_childcare_benefits_scandal.

Wikipedia. “Electronic Commerce Directive 2000”. https://en.wikipedia.org/w/index.php?title=Electronic_Commerce_Directive_2000&oldid=1315198928.

Wikipedia. “ELIZA”. <https://en.wikipedia.org/w/index.php?title=ELIZA&oldid=1318801380>.

Wikipedia. “Futurama”. <https://en.wikipedia.org/w/index.php?title=Futurama&oldid=1318939071>.

Wikipedia. “Gartner Hype Cycle”. https://en.wikipedia.org/w/index.php?title=Gartner_hype_cycle&oldid=1305483972.

Wikipedia. “Gradient Descent”, 2 November 2025, https://en.wikipedia.org/w/index.php?title=Gradient_descent&oldid=1320054958.

Wikipedia. “Hallucination (Artificial Intelligence)”. [https://en.wikipedia.org/w/index.php?title=Hallucination_\(artificial_intelligence\)&oldid=1320533332](https://en.wikipedia.org/w/index.php?title=Hallucination_(artificial_intelligence)&oldid=1320533332).

Wikipedia. “Information Superhighway”. https://en.wikipedia.org/w/index.php?title=Information_superhighway&oldid=1306198017.

Wikipedia. “Large language model”. https://en.wikipedia.org/wiki/Large_language_model.

Wikipedia. “MacPherson v. Buick Motor”. https://en.wikipedia.org/w/index.php?title=MacPherson_v._Buick_Motor_Co.&oldid=1311379098.

Wikipedia. “Machine Intelligence Research Institute”. https://en.wikipedia.org/wiki/Machine_Intelligence_Research_Institute.

Wikipedia. “Manifest Destiny”. https://en.wikipedia.org/w/index.php?title=Manifest_destiny&oldid=1316761617.

Wikipedia. “Mycin”. <https://en.wikipedia.org/wiki/Mycin>.

- Wikipedia. “Nirvana Fallacy”. https://en.wikipedia.org/w/index.php?title=Nirvana_fallacy&oldid=1307267896.
- Wikipedia. “P(doom)”. [https://en.wikipedia.org/w/index.php?title=P\(doom\)&oldid=1319821713](https://en.wikipedia.org/w/index.php?title=P(doom)&oldid=1319821713).
- Wikipedia. “Prosperity Theology”. https://en.wikipedia.org/wiki/Prosperity_theology.
- Wikipedia. “Reward Hacking”. https://en.wikipedia.org/w/index.php?title=Reward_hacking&oldid=1319181065.
- Wikipedia. “Robodebt scheme”. https://en.wikipedia.org/wiki/Robodebt_scheme.
- Wikipedia. “Roko’s basilisk”. https://en.wikipedia.org/wiki/Roko’s_basilisk.
- Wikipedia. “Rote Learning”. https://en.wikipedia.org/w/index.php?title=Rote_learning&oldid=1299277683.
- Wikipedia. “Roy Amara”. https://en.wikipedia.org/w/index.php?title=Roy_Amara&oldid=1306707783.
- Wikipedia. “Scapegoat”. <https://en.wikipedia.org/w/index.php?title=Scapegoat&oldid=1316414935>.
- Wikipedia. “Semantic Gap”. https://en.wikipedia.org/w/index.php?title=Semantic_gap&oldid=1312781776.
- Wikipedia. “Sidewalk Labs”. https://en.wikipedia.org/wiki/Sidewalk_Labs.
- Wikipedia, “Sophia (robot)”. [https://en.wikipedia.org/w/index.php?title=Sophia_\(robot\)&oldid=1316421104](https://en.wikipedia.org/w/index.php?title=Sophia_(robot)&oldid=1316421104).
- Wikipedia, “White-Smith Music Publishing Co. v. Apollo Co”. https://en.wikipedia.org/wiki/White-Smith_Music_Publishing_Co._v._Apollo_Co.
- Wilkins, Joe. “World’s First AI Minister Is ‘Pregnant’ With 83 Offspring Government Announces”. *Futurism*, 28 October 2025. <https://futurism.com/artificial-intelligence/rama-diella-albania-pregnant>.
- Williamson, Ben. *Big Data in Education: The Digital Future of Learning, Policy and Practice*. SAGE Publications Ltd, 2017. <https://us.sagepub.com/en-us/nam/big-data-in-education/book249086>.
- Willison, Simon. “The Surprise Deprecation of GPT-4o for ChatGPT Consumers”. *Simon Willison’s Weblog*, 8 August 2025. <https://simonwillison.net/2025/Aug/8/surprise-deprecation-of-gpt-4o/>.

Wilson, James Q. *Bureaucracy: What Government Agencies Do and Why They Do It*. Basic Books, 1989. <http://archive.org/details/bureaucracywhatg00wils>.

Wilson, Mark. “AI Is Inventing Languages Humans Can’t Understand. Should We Stop It?” *Fast Company*, 14 July 2017. <https://www.fastcompany.com>.

Windwehr, Svea, Jillian C. York, and Christoph Schmon. “Just Banning Minors From Social Media Is Not Protecting Them”. *Electronic Frontier Foundation*, 28 July 2025. <https://www.eff.org/deeplinks/2025/07/just-banning-minors-social-media-not-protecting-them>.

Winner, Langdon. *Autonomous Technology: Technics-Out-of-Control as a Theme in Political Thought*. MIT Press, 1977. <https://mitpress.mit.edu/9780262230780/autonomous-technology/>.

Winner, Langdon. “Do Artifacts Have Politics?” *Daedalus* 109, no. 1 (1980): 121–36.

Woollacott, Emma. “What Is ‘AI Washing’ and Why Is It a Problem?” *BBC*, 26 June 2024. <https://www.bbc.com/news/articles/c9xx8122893o>.

Wu, Tim. “A Brief History of American Telecommunications Regulation”. *Oxford International Encyclopedia Of Legal History* 5 (2009): 95. https://scholarship.law.columbia.edu/faculty_scholarship/1461/.

Wu, Tim. “The Master Switch: The Rise and Fall of Information Empires”. *Columbia Law School*, 2007. https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=2462&context=faculty_scholarship.

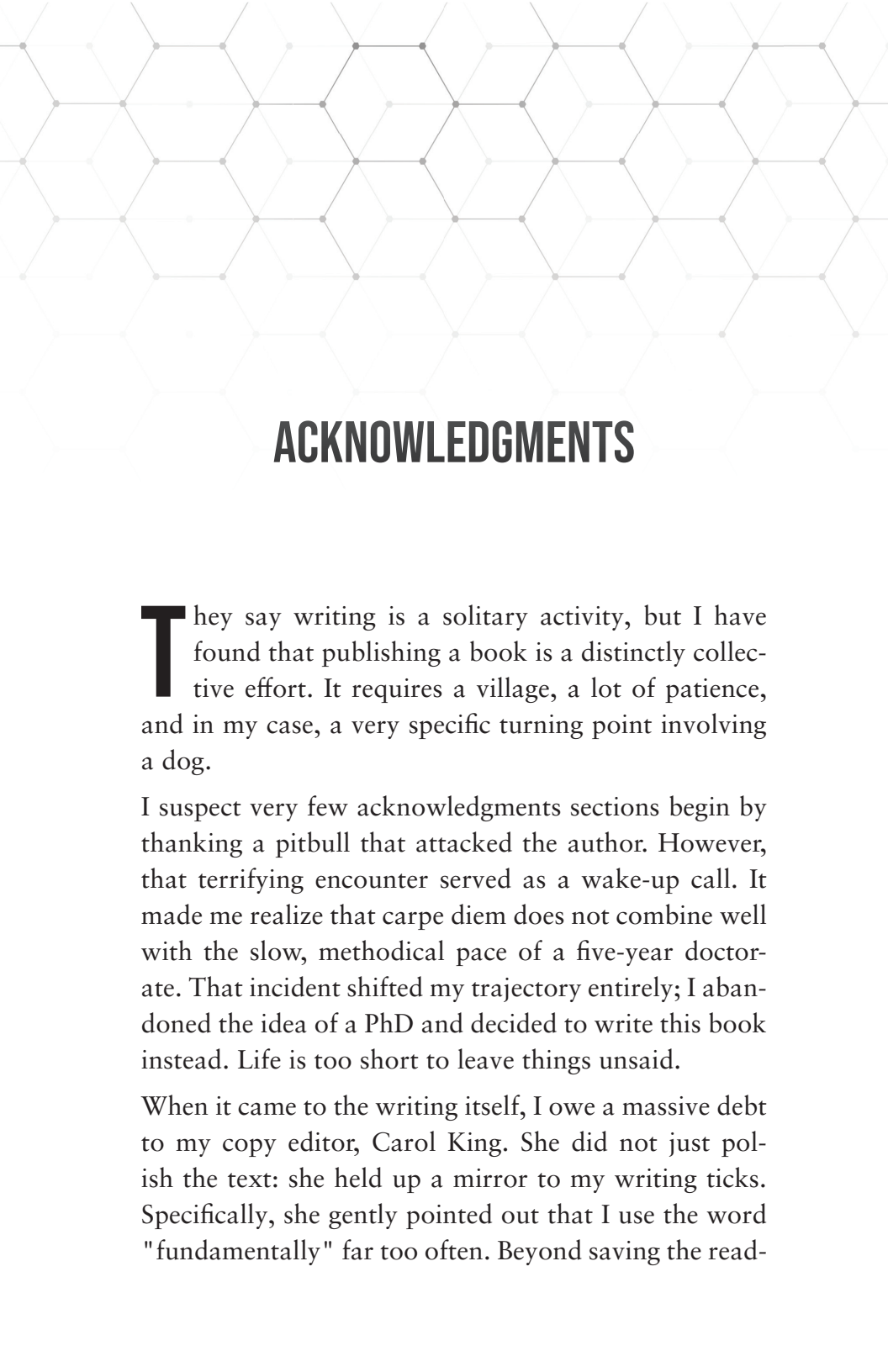
Yerushalmy, Jonathan. “‘I Want to Destroy Whatever I Want’: Bing’s AI Chatbot Unsettles US Reporter”. *Technology. The Guardian*, 17 February 2023. <https://www.theguardian.com/technology/2023/feb/17/i-want-to-destroy-whatever-i-want-bings-ai-chatbot-unsettles-us-reporter>.

Yin, Ziqi, Hao Wang, Kaito Horio, Daisuke Kawahara, and Satoshi Sekine. “Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance”. *arXiv:2402.14531*. Preprint, arXiv, 14 October 2024. <https://doi.org/10.48550/arXiv.2402.14531>.

York, Jillian C. *Silicon Values: The Future of Free Speech Under Surveillance Capitalism*. Verso Books, 2021. <https://www.versobooks.com/en-gb/products/882-silicon-values>.

Zhang, Yutong, Dora Zhao, Jeffery T. Hancock, Robert Kraut, and Diyi Yang. *The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being*. Arxiv. 2025. <https://arxiv.org/html/2506.12605v1#bib.bib100>.

Zuckerberg, Mark. "Personal Superintelligence". Meta, 18 January 2024. <https://www.meta.com/superintelligence/>.



ACKNOWLEDGMENTS

They say writing is a solitary activity, but I have found that publishing a book is a distinctly collective effort. It requires a village, a lot of patience, and in my case, a very specific turning point involving a dog.

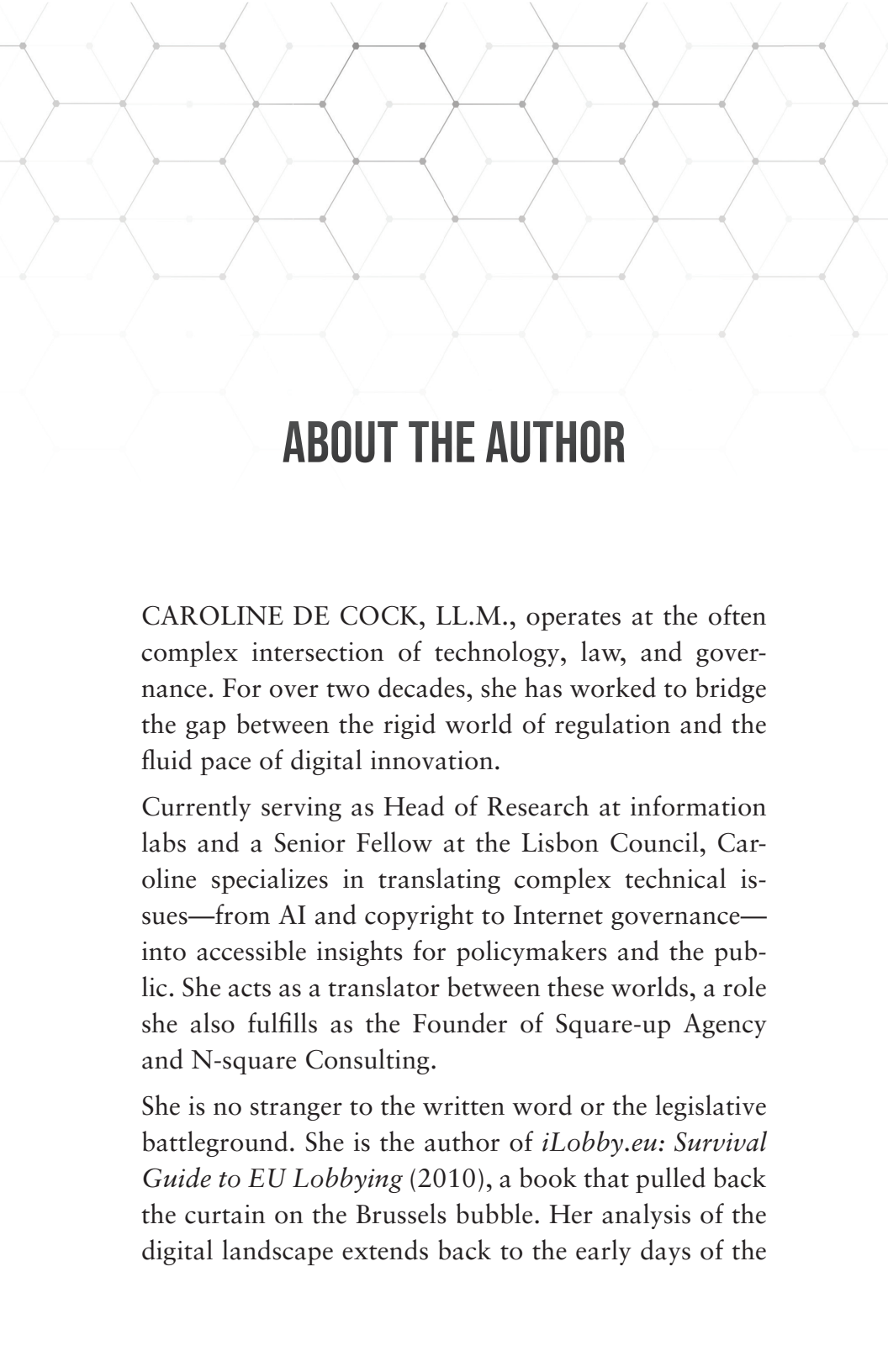
I suspect very few acknowledgments sections begin by thanking a pitbull that attacked the author. However, that terrifying encounter served as a wake-up call. It made me realize that *carpe diem* does not combine well with the slow, methodical pace of a five-year doctorate. That incident shifted my trajectory entirely; I abandoned the idea of a PhD and decided to write this book instead. Life is too short to leave things unsaid.

When it came to the writing itself, I owe a massive debt to my copy editor, Carol King. She did not just polish the text: she held up a mirror to my writing ticks. Specifically, she gently pointed out that I use the word "fundamentally" far too often. Beyond saving the read-

er from my repetitive vocabulary, she was instrumental in the book's evolution. She pushed me to ensure the journey from the first draft to the final version resulted in something crisper. She insisted the text remain action-focused, preventing me from getting lost in the weeds of theory.

I am equally grateful to Glyn Moody, who bravely accepted the task of the ultimate proofread. His constructive feedback on the final manuscript was invaluable, helping me spot any remaining cracks in the foundation and giving me the confidence to cross the finish line.

Finally, ideas do not exist in a vacuum. I must thank Luciano Floridi and Anna Nobre. Their article on the anthropomorphization of AI was a lightbulb moment for me. It provided the precise language I needed to articulate my lingering frustrations regarding the interaction between policy and technology. Their work gave me the tools to express what I had been feeling: that we must stop treating our creations like gods.



ABOUT THE AUTHOR

CAROLINE DE COCK, LL.M., operates at the often complex intersection of technology, law, and governance. For over two decades, she has worked to bridge the gap between the rigid world of regulation and the fluid pace of digital innovation.

Currently serving as Head of Research at information labs and a Senior Fellow at the Lisbon Council, Caroline specializes in translating complex technical issues—from AI and copyright to Internet governance—into accessible insights for policymakers and the public. She acts as a translator between these worlds, a role she also fulfills as the Founder of Square-up Agency and N-square Consulting.

She is no stranger to the written word or the legislative battleground. She is the author of *iLobby.eu: Survival Guide to EU Lobbying* (2010), a book that pulled back the curtain on the Brussels bubble. Her analysis of the digital landscape extends back to the early days of the

internet economy, with contributions as a researcher to *The New Virtual Money and Business and Law on the Internet*.

Beyond books, Caroline is a prolific commentator, blogger, and podcaster who actively shapes the debate on digital rights. Her writing and analysis have appeared in *Techdirt*, *Euractiv*, and *Techpolicy*, among others.

Her perspective is shaped by a life spent crossing borders. Having lived in eleven countries across four continents—from Africa to South America—she brings a deeply global lens to the challenges of digital regulation. She understands that policy does not exist in a vacuum, and that the metaphors we use to describe technology can have real-world consequences. This focus on narrative, clarity, and democratic accountability anchors her latest work, *AI Tools, Not Gods*.